

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2005, 10:25:11 ; Search time 161 Seconds
(without alignments)
67.263 Million cell updates/sec

Title: US-10-083-576-1

Perfect score: 137

Sequence: 1 LKQLDAQQTQLRIDSFFRLAQEKEDKR 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A. Geneseq_15Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	128	93.4	29	ADI47228	Cancer-sp
2	128	93.4	550	ADE52960	FEN-1 rel
3	128	93.4	551	ADA66102	DNAP-rela
4	128	93.4	1186	ADP12557	Protein e
5	119	86.9	135	AAO02970	Human pol
6	117.5	85.8	30	AA884854	XPG pepti
7	110	80.3	543	ADE52961	FEN-1 rel
8	110	80.3	543	ADA66103	DNAP-rela
9	96	70.1	527	ADE52962	FEN-1 rel
10	96	70.1	527	ADA66104	DNAP-rela
11	57	41.6	493	ABR53625	Protein s
12	57	41.6	493	ADK64206	Disease t
13	56	40.9	1965	ADN20626	Disease t
14	53.5	39.1	730	ABB71879	Drosophil
15	52	38.0	978	ABU16227	Protein e
16	52	38.0	981	AAU36618	Staphyloc
17	52	38.0	981	ABM71093	Staphyloc
18	51	37.2	336	ABP79659	N. gonorr
19	50	36.5	933	ABB60762	Drosophil
20	49	35.8	288	ABO81220	Pseudomon
21	49	35.8	474	AA34722	Human sec
22	49	35.8	625	ADR10287	Human pro
23	49	35.8	958	AA51120	Human SAR
24	49	35.8	958	ADL83102	Human PRO
25	48	35.0	412	ABU33976	Protein e

26	47	34.3	249	8	ADO61961	Ado61961 Transcrip
27	47	34.3	385	4	ABB63960	Abb63960 Drosophil
28	47	34.3	478	7	ADF04901	Adf04901 Bacterial
29	47	34.3	518	3	AA553322	AA553322 Human col
30	47	34.3	555	4	AA875073	AA875073 Human col
31	47	34.3	667	5	ABB05040	Abb05040 Human NOV
32	47	34.3	667	5	ABB05041	Abb05041 Human NOV
33	47	34.3	671	3	AA599426	AA599426 Human PRO
34	47	34.3	671	4	AA661175	AA661175 Protein o
35	47	34.3	671	4	AAU29196	Aau29196 Human PRO
36	47	34.3	671	5	ABB05042	Abb05042 Human NOV
37	47	34.3	671	5	AAU97038	Aau97038 Human LP
38	47	34.3	671	6	ABU58572	Abu58572 Human PRO
39	47	34.3	671	6	ABU88120	Abu88120 Novel hum
40	47	34.3	671	6	ABU84435	Abu84435 Human sec
41	47	34.3	671	6	ABR66309	ABr66309 Human sec
42	47	34.3	671	6	ABR65699	ABr65699 Human sec
43	47	34.3	671	6	ABU99639	Abu99639 Human sec
44	47	34.3	671	6	ABU82878	Abu82878 Human PRO
45	47	34.3	671	6	ABU89999	Abu89999 Novel hum

ALIGNMENTS

RESULT 1
ADI47228
ID ADI47228 standard; peptide; 29 AA.
XX
AC ADI47228;
XX
DT 06-MAY-2004 (first entry)
XX
DE Cancer-specific proliferating cell nuclear antigen binding XPG peptide.
XX
KW cancer-specific proliferating cell nuclear antigen; cspcna; immunoassay;
KW indicator; cancer; XPG.
XX
OS Unidentified.
XX
PN US2003162233-A1.
XX
PD 28-AUG-2003.
XX
PF 27-FEB-2002; 2002US-00083576.
XX
PR 27-FEB-2002; 2002US-00083576.
XX
PA (MALK/) MALKAS L H.
PA (HICK/) HICKEY R J.
PA (BECH/) BECHTEL P E.
PA (SCHN/) SCHNAPER L.
PA (PARK/) PARK M.
PA (HOEL/) HOELZ D J.
PA (TOMI/) TOMIC D.
XX
Malkas LH, Hickey RJ, Bechtel PE, Schnaper L, Park M, Hoelz DJ, Tomic D;
XX
WPI; 2003-766312/72.
XX
Purification of cancer-specific proliferating cell nuclear antigen, useful for producing antibodies, from a tissue or body fluid sample comprises contacting the sample with an immobilized peptide.
XX
Claim 1; SEQ ID NO 1; 20pp; English.
XX
The invention relates to a novel method for purifying a cancer-specific proliferating cell nuclear antigen (cspcna) from a tissue or body fluid sample. The novel method comprises contacting the sample with a peptide immobilized on a solid support and isolating the cspcna from the resulting peptide-cspcna complex. The purified cspcna is useful for producing antibodies, e.g. for use in immunoassays, as cspcna is a good

CC diagnostic/prognostic indicator for cancer. This sequence represents the
 CC peptide for contacting with the sample on a solid support used in the
 CC method of the invention.

CC Sequence 29 AA;

Query Match 93.4%; Score 128; DB 7; Length 29;
 Best Local Similarity 92.9%; Pred. No. 1.1e-11;
 Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LKQIDAQQTOLRIDSPFRLAQKEKDKR 28
 |||||
 Db 1 LKQIDAQQTOLRIDSPFRLAQKEKEDAK 28

RESULT 2
 ADE52960
 ID ADE52960 standard; protein; 550 AA.

XX ADE52960;

DT 29-JAN-2004 (first entry)

XX FEN-1 related polypeptide used within the scope of the invention, #15.

XX Flap endonuclease-1; FEN-1; endonuclease; structure-specific nuclease;
 KW invasive cleavage structure; thermostable; DNA polymerase; 5' nuclease;
 KW viral infection; bacterial infection; cancer; forensic analysis;
 KW paternity determination.

XX Homo sapiens.

XX WO200270755-A2.

XX 12-SEP-2002.

XX 15-NOV-2001; 2001WO-US044953.

XX 15-NOV-2000; 2000US-00713601.

XX 17-NOV-2000; 2000US-00714935.

XX (THIR-) THIRD WAVE TECHNOLOGIES INC.

XX Lyamichev VI, Kaiser MW, Lyamicheva N;

XX WPI; 2002-750464/81.

XX New composition useful for detecting and characterizing nucleic acid
 PT sequences and sequence variants for detecting the presence of viral or
 PT bacterial infections or cancer, comprises purified or chimerical FEN-1
 PT endonuclease.

XX Disclosure; SEQ ID NO 142; 871pp; English.

XX The invention discloses a new composition (I) which comprises a purified
 CC Flap endonuclease-1 (FEN-1) from e.g. *Sulfolobus solfataricus*,
 CC Pyrobaculum aerophilum or a chimerical FEN-1 endonuclease having a
 CC portion of the above endonuclease in addition to that of *Pyrococcus*
 CC horikoshii and *Aeropyrum pernix*. Also claimed is a composition comprising
 CC an isolated nucleic acid sequence encoding the endonuclease mentioned
 CC above, a composition comprising a vector having the nucleic acid sequence
 CC cited above, a composition comprising a host cell and vector cited above,
 CC a mixture comprising a first structure-specific nuclease selected from
 CC the species mentioned in composition (I), and a purified second structure
 CC -specific nuclease and detecting a target sequence, comprising: (a)
 CC providing a sample suspected of containing the target sequence,
 CC oligonucleotides capable of forming an invasive cleavage structure in the
 CC presence of the target sequence, and a FEN-1 endonuclease selected from
 CC the species cited above and (b) exposing the sample to the
 CC oligonucleotides and FEN-1 endonuclease. The second structure-specific
 CC nuclease also comprises a thermostable DNA polymerase. It has a 5'
 CC nuclease derived from a DNA polymerase altered in amino acid sequence
 CC such that it exhibits reduced DNA synthetic activity from that of the

CC wild-type DNA polymerase but retains substantially the same 5' nuclease
 CC activity of the wild-type DNA polymerase. The second structure is
 CC selected from CLEAVASE BN enzyme, CLEAVASE DA enzyme, CLEAVASE DN enzyme,
 CC CLEAVASE DV enzyme, CLEAVASE BN/thrombin enzyme, CLEAVASE TthDN enzyme,
 CC T. aquaticus DNA polymerase, T. thermophilus DNA polymerase, E. coli Exo
 CC III and S. cerevisiae Rad1/Rad10 complex. The nucleic acid treatment kit
 CC comprises (I) and oligonucleotides capable of forming an invasive
 CC cleavage structure in the presence of a target nucleic acid. The
 CC oligonucleotides comprise: (a) a first oligonucleotide having a 5'
 CC portion complementary to a first portion of a target nucleic acid and (b)
 CC a second oligonucleotide comprising a 5' portion complementary to a
 CC second portion of the target nucleic acid downstream of and contiguous to
 CC the first portion and a 3' portion. The 3' portion of the second
 CC oligonucleotide comprises a single 3' terminal nucleotide not
 CC complementary to the target nucleic acid. Additionally, the kit has a
 CC third oligonucleotide complementary to a third portion of the target
 CC nucleic acid upstream of the first portion of the first target nucleic
 CC acid. In detecting a target sequence, the oligonucleotides and
 CC endonuclease are mixed under conditions where an invasive cleavage
 CC structure is formed between the target sequence and the oligonucleotides
 CC if the target sequence is present in the sample, where the invasive
 CC cleavage structure is cleaved by the endonuclease to form a cleavage
 CC product. The composition is useful in detecting and characterizing
 CC specific nucleic acid sequences and sequence variants which can be used
 CC in detecting the presence of viral or bacterial infections, and other
 CC diseases such as cancer. The composition may also be used in forensic
 CC analysis or for paternity determinations. The sequence presented is a FEN
 CC -1 related polypeptide used within the scope of the invention.

XX Sequence 550 AA;

Query Match 93.4%; Score 128; DB 5; Length 550;
 Best Local Similarity 92.9%; Pred. No. 2.5e-10;
 Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LKQIDAQQTOLRIDSPFRLAQKEKDKR 28
 |||||
 Db 345 LKQIDAQQTOLRIDSPFRLAQKEKEDAK 372

RESULT 3
 ADA66102

ID ADA66102 standard; protein; 551 AA.

XX ADA66102;

DT 20-NOV-2003 (first entry)

XX DNAP-related protein #8.

XX DNAP; invasive cleavage structure; dendrimer; nuclease; endonuclease;
 KW polymerase; polyglycol; 5' nuclease; allelic variation.

XX Homo sapiens.

XX US2003044796-A1.

XX 06-MAR-2003.

XX 27-AUG-2001; 2001US-00940244.

XX 26-NOV-1996; 96US-00756386.

XX 24-MAR-1998; 98WO-US005809.

XX 09-JUL-1999; 99US-00350309.

XX 08-FEB-2000; 2000US-00381212.

XX 08-DEC-2000; 2000US-00732622.

XX (NERI/) NERI B P.

XX (HALL/) HALL J G.

XX (LYAM/) LYAMICHEV V.

XX (SMIT/) SMITH L M.

PI Neri BP, Hall JG, Lyamichev V, Smith LM;

XX WPI; 2003-596420/56.
 XX Detection system for nucleic acid sequences comprises oligonucleotides
 PT configured for hybridizing to target nucleic acid to form invasive
 PT cleavage structure and dendrimer.
 XX
 XX Disclosure; Fig 59; 354pp; English.
 XX The invention relates to a detection system which has oligonucleotides
 CC configured for hybridisation to a target nucleic acid to form an invasive
 CC cleavage structure and dendrimer, where the oligonucleotides are attached
 CC to the dendrimer. The invention also relates to a method for
 CC characterising a nucleic acid sequence comprising providing a sample
 CC suspected of containing a target nucleic acid, oligonucleotides
 CC configured to hybridise to the target nucleic acid to form an invasive
 CC cleavage structure and a dendrimer to which the oligonucleotide is
 CC attached, and exposing the sample to the oligonucleotides and an agent
 CC that detects the presence of an invasive cleavage structure. The agent
 CC comprises a cleavage agent having a structure-specific nuclease,
 CC preferably a 5' nuclease comprising an endonuclease or polymerase. The
 CC detection system further comprises a spacer molecule, consisting of a
 CC carbon chain, polynucleotide or polyglycol, to which the oligonucleotide
 CC is attached. The target molecule and the agent are attached to a solid
 CC support. The invention is used in the detection and characterisation of
 CC nucleic acid sequences and variations in these sequences, used in allelic
 CC variation studies. This sequence represents a protein used in the scope
 CC of the invention.
 XX
 XX Sequence 551 AA;

Query Match 93.4%; Score 128; DB 7; Length 551;
 Best Local Similarity 92.9%; Pred. No. 2.5e-10;
 Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LKQLDAQQTQLRIDSFFRLAQOQEKDKR 28
 |||||
 Db 345 LKQLDAQQTQLRIDSFFRLAQOQEKDAK 372

RESULT 4
 ADP12557
 ID ADP12557 standard; protein; 1186 AA.
 XX
 AC ADP12557;
 XX
 DT 12-AUG-2004 (first entry)
 XX
 DE Protein encoded by mRNA of the invention #167.
 XX
 KW transplant rejection; immune system; rheumatoid arthritis; lupus;
 KW inflammatory bowel disease; multiple sclerosis; HIV; AIDS.
 XX
 OS Homo sapiens.
 XX
 PN WO2004042346-A2.
 XX
 PD 21-MAY-2004.
 XX
 PF 24-APR-2003; 2003WO-US012946.
 XX
 XX 24-APR-2002; 2002US-00131831.
 PR 20-DEC-2002; 2002US-00325899.
 XX
 XX (EXPR-) EXPRESSION DIAGNOSTICS INC.
 PA
 XX Wohlgenuth J, Fry K, Woodward R, Ly N, Prentice J, Morris M;
 PI Rosenberg S;
 PI WPI; 2004-400724/37.
 DR
 XX Diagnosing or monitoring transplant rejection, e.g. heart, kidney, liver,
 PT pancreas, pancreatic islet, lung, bone marrow or stem cell transplant

PT rejection, in an individual, comprises detecting the expression level of
 PT the genes.
 XX
 XX Claim 65; SEQ ID NO 2566; 1762pp; English.
 XX
 CC The present invention relates to diagnosing or monitoring transplant
 CC rejection, e.g. cardiac or kidney transplant rejection, in an individual
 CC comprises detecting the expression level of one or more genes. The
 CC methods, system and kits are useful in diagnosing or monitoring
 CC transplant rejection, e.g. heart, kidney, liver, pancreas, pancreatic
 CC islet, lung, bone marrow or stem cell transplant rejection, in an
 CC xenotransplant rejection or mechanical organ replacement rejection, in an
 CC individual. The method is also useful in assessing the immune status of
 CC an individual. The methods are also useful in diagnosing and monitoring
 CC diseases that involve the immune system, e.g. rheumatoid arthritis,
 CC lupus, inflammatory bowel diseases, multiple sclerosis, HIV/AIDS or
 CC viral, bacterial or fungal infection. The present sequence represents a
 CC protein that is encoded by the mRNA of the invention.
 XX
 XX Sequence 1186 AA;

Query Match 93.4%; Score 128; DB 8; Length 1186;
 Best Local Similarity 92.9%; Pred. No. 5.8e-10;
 Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LKQLDAQQTQLRIDSFFRLAQOQEKDKR 28
 |||||
 Db 981 LKQLDAQQTQLRIDSFFRLAQOQEKDAK 1008

RESULT 5
 AAO02970
 ID AAO02970 standard; protein; 135 AA.
 XX
 AC AAO02970;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 16862.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US004927.
 XX
 PR 28-FEB-2000; 2000US-00515126.
 PR 18-MAY-2000; 2000US-00577409.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Drmanac RT;
 PI WPI; 2001-514838/56.
 DR N-PSDB; AAI82901.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
 PT and treating e.g. leukemia, inflammation and immune disorders.
 XX
 XX Claim 20; SEQ ID NO 16862; 1399pp + Sequence Listing; English.
 PS
 XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX SQ Sequence 135 AA;

Query Match 86.9%; Score 119; DB 4; Length 135;
 Best Local Similarity 85.7%; Pred. No. 1.2e-09;
 Matches 24; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LKQLDAQQTQLRIDSFFRLAQOQEKDKR 28
 Db 52 LKQLDAQQTQLRIDSFFRLAQOQEKDKR 79

RESULT 6
 ID AAB84854
 AC AAB84854 standard; peptide; 30 AA.

XX AAB84854;

XX 30-JUL-2001 (first entry)

XX XPG peptide fragment.

XX p15-PAF cell cycle protein; R0101; mitosis; cell cycle;
 XX proliferating cell nuclear antigen; PCNA; cancer; XPG.

XX Unidentified.

XX Key Location/Qualifiers
 XX 11-18
 XX /label= PCNA-binding_motif

XX WO200129072-A2.

XX 26-APR-2001.

XX 18-OCT-2000; 2000WO-US041296.

XX 18-OCT-1999; 99US-00420092.

XX (RIGE-) RIGEL PHARM INC.

XX Luo Y, Yu P, Huang B;

XX WPI; 2001-290894/30.

XX Screening modulator of p15PAF cell cycle protein R0101 activity, by
 XX adding a candidate bioactive agent to cell comprising recombinant nucleic
 XX acid encoding protein, and determining the agents effect.

XX Disclosure; Fig 2B; 67pp; English.

XX The present invention relates to a method for screening for a bioactive
 XX agent capable of modulating the activity of p15-PAF cell cycle protein
 XX (also named R0101). p15-PAF is involved in mitosis and the cell cycle and
 XX binds to Proliferating Cell Nuclear Antigen (PCNA). Agents identified by
 XX the present method are useful for the treatment or immunotherapy of p15-
 XX PAF cell cycle protein related disorders e.g. cancers. The present
 XX sequence is a peptide fragment of XPG protein. This sequence contains a
 XX PCNA-binding motif, and this sequence was used in a sequence homology
 XX alignment with other PCNA-binding motif containing proteins

XX SQ Sequence 30 AA;

Query Match 85.8%; Score 117.5; DB 4; Length 30;
 Best Local Similarity 89.7%; Pred. No. 4e-10;

Matches 26; Conservative 1; Mismatches 1; Indels 1; Gaps 1;
 Qy 1 LKQLDA-QQTQLRIDSFFRLAQOQEKDKR 28
 Db 1 LKQLDAQQTQLRIDSFFRLAQOQEKDKR 29

RESULT 7

ADES2961
 ID ADE52961 standard; protein; 543 AA.

XX ADE52961;

XX 29-JAN-2004 (first entry)

XX FEN-1 related polypeptide used within the scope of the invention, #16.
 XX Flap endonuclease-1; FEN-1; endonuclease; structure-specific nuclease;
 XX invasive cleavage structure; thermostable; DNA polymerase; 5' nuclease;
 XX viral infection; bacterial infection; cancer; forensic analysis;
 XX paternity determination.

XX Mus musculus.

XX WO200270755-A2.

XX 12-SEP-2002.

XX 15-NOV-2001; 2001WO-US044953.

XX 15-NOV-2000; 2000US-00713601.

XX 17-NOV-2000; 2000US-00714935.

XX (THIR-) THIRD WAVE TECHNOLOGIES INC.

XX Lyamichev VI, Kaiser MW, Lyamicheva N;

XX WPI; 2002-750464/81.

XX New composition useful for detecting and characterizing nucleic acid
 XX sequences and sequence variants for detecting the presence of viral or
 XX bacterial infections or cancer, comprises purified or chimerical FEN-1
 XX endonuclease.

XX Disclosure; SEQ ID NO 143; 871pp; English.

XX The invention discloses a new composition (I) which comprises a purified
 XX flap endonuclease-1 (FEN-1) from e.g. Sulfolobus solfataricus,
 XX Pyrobaculum aerophilum or a chimerical FEN-1 endonuclease having a
 XX portion of the above endonuclease in addition to that of Pyrococcus
 XX horikoshii and Aeropyrum pernix. Also claimed is a composition comprising
 XX an isolated nucleic acid sequence encoding the endonuclease mentioned
 XX above, a composition comprising a vector having the nucleic acid sequence
 XX cited above, a composition comprising a host cell and vector cited above,
 XX a mixture comprising a first structure-specific nuclease selected from
 XX the species mentioned in composition (I), and a purified second structure
 XX -specific nuclease and detecting a target sequence, comprising: (a)
 XX providing a sample suspected of containing the target sequence,
 XX oligonucleotides capable of forming an invasive cleavage structure in the
 XX presence of the target sequence, and a FEN-1 endonuclease selected from
 XX the species cited above and (b) exposing the sample to the
 XX oligonucleotides and FEN-1 endonuclease. The second structure-specific
 XX nuclease also comprises a thermostable DNA polymerase. It has a 5'
 XX nuclease derived from a DNA polymerase altered in amino acid sequence
 XX such that it exhibits reduced DNA synthetic activity from that of the
 XX wild-type DNA polymerase but retains substantially the same 5' nuclease
 XX activity of the wild-type DNA polymerase. The second structure is
 XX selected from CLEAVASE BN enzyme, CLEAVASE DA enzyme, CLEAVASE DN enzyme,
 XX CLEAVASE DV enzyme, CLEAVASE BN/thrombin enzyme, CLEAVASE TthDN enzyme,
 XX T. aquaticus DNA polymerase, T. thermophilus DNA polymerase, E. coli Exo
 XX III and S. cerevisiae Radi/Radio complex. The nucleic acid treatment kit
 XX comprises (I) and oligonucleotides capable of forming an invasive
 XX cleavage structure in the presence of a target nucleic acid. The

XX The invention relates to a detection system which has oligonucleotides
CC configured for hybridisation to a target nucleic acid to form an invasive
CC cleavage structure and dendrimer, where the oligonucleotides are attached
CC to the dendrimer. The invention also relates to a method for
CC characterising a nucleic acid sequence comprising providing a sample
CC suspected of containing a target nucleic acid, oligonucleotides
CC configured to hybridise to the target nucleic acid to form an invasive
CC cleavage structure and a dendrimer to which the oligonucleotide is
CC attached, and exposing the sample to the oligonucleotides and an agent
CC that detects the presence of an invasive cleavage structure. The agent
CC comprises a cleavage agent having a structure-specific nuclease,
CC preferably a 5' nuclease comprising an endonuclease or polymerase. The
CC detection system further comprises a spacer molecule, consisting of a
CC carbon chain, polynucleotide or polyglycol, to which the oligonucleotide
CC is attached. The target molecule and the agent are attached to a solid
CC support. The invention is used in the detection and characterisation of
CC nucleic acid sequences and variations in these sequences, used in allelic
CC variation studies. This sequence represents a protein used in the scope
CC of the invention.

Query Match 80.3%; Score 110; DB 7; Length 543;
Best Local Similarity 78.6%; Pred. No. 1.2e-07;
Matches 22; Conservative 3; Mismatches 3; Indels 0; Gaps 0

Qy 1 LKQLDAQOTQLRIDSFRLAQOEKDKR 28
 ||| : ||| ||| ||| ||| ||| ||| :
Dd 345 LXHLNAHOTQLRIDSFRLAQOEKODAK 372

RESULT 9	
ADE52962	
ID	ADE52962 standard; protein; 527 AA.
XX	
XX	AC ADE52962;
XX	
XX	
DT	29-JAN-2004 (first entry)
XX	
XX	
DE	FEN-1 related polypeptide used within the scope of the invention, #17.
XX	
KW	Flap endonuclease-1; FEN-1; endonuclease; structure-specific nuclease
KW	invasive cleavage structure; thermostable; DNA polymerase; 5' nucleases
KW	viral infection; bacterial infection; cancer; forensic analysis;
KW	paternity determination.

KW viral infection, bacteri
paternity determination,
KW

Xenopus laevis.

PN WQ200270755-A2

12-SEP-2002

15-NOV-2001 2001W0-1150449E3

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

PR 17-NOV-2000; 2000US-00714935.

PA (THIR-) THIRD WAVE TECHNOLOGIES INC.

PI Lyamichev VI, Kaiser MW, Lyamicheva N;

DR WPI; 2002-750464/81.

PT New composition useful for detecting and characterizing nucleic acid

PT bacterial infections or cancer, comprises purified or chimerical FEN-1

XX
XX
PS Disclosure; SEQ ID NO 144; 871pp; English.
XX
XX

flap endonuclease-1 (FEN-1) from e.g. *Sulfolobus solfataricus*,

CC Pyrobaculum aerophilum or a chimerical FEN-1 endonuclease having a
 CC portion of the above endonuclease in addition to that of Pyrococcus
 CC horikoshii and Aeropyrum pernix. Also claimed is a composition comprising
 CC an isolated nucleic acid sequence encoding the endonuclease mentioned
 CC above, a composition comprising a vector having the nucleic acid sequence
 CC cited above, a composition comprising a host cell and vector cited above,
 CC a mixture comprising a first structure-specific nuclease selected from
 CC the species mentioned in composition (I), and a purified second structure
 CC -specific nuclease and detecting a target sequence, comprising: (a)
 CC providing a sample suspected of containing the target sequence,
 CC oligonucleotides capable of forming an invasive cleavage structure in the
 CC presence of the target sequence, and a FEN-1 endonuclease selected from
 CC the species cited above and (b) exposing the sample to the
 CC oligonucleotides and FEN-1 endonuclease. The second structure-specific
 CC nuclease also comprises a thermostable DNA polymerase. It has a 5'
 CC nuclease derived from a DNA polymerase altered in amino acid sequence
 CC such that it exhibits reduced DNA synthetic activity from that of the
 CC wild-type DNA polymerase but retains substantially the same 5' nuclease
 CC activity of the wild-type DNA polymerase. The second structure is
 CC selected from CLEAVASE BN enzyme, CLEAVASE DA enzyme, CLEAVASE DN enzyme,
 CC CLEAVASE DV enzyme, CLEAVASE BN/thrombin enzyme, CLEAVASE TthDN enzyme,
 CC T. aquaticus DNA polymerase, T. thermophilus DNA polymerase, E. coli Exo
 CC III and S. cerevisiae Rad1/Rad10 complex. The nucleic acid treatment kit
 CC comprises (I) and oligonucleotides capable of forming an invasive
 CC cleavage structure in the presence of a target nucleic acid. The
 CC oligonucleotides comprise: (a) a first oligonucleotide having a 5'
 CC portion complementary to a first portion of a target nucleic acid and (b)
 CC a second oligonucleotide comprising a 5' portion complementary to a
 CC second portion of the target nucleic acid downstream of and contiguous to
 CC the first portion and a 3' portion. The 3' portion of the second
 CC oligonucleotide comprises a single 3' terminal nucleotide not
 CC complementary to the target nucleic acid. Additionally, the kit has a
 CC third oligonucleotide complementary to a third portion of the target
 CC nucleic acid upstream of the first portion of the first target nucleic
 CC acid. In detecting a target sequence, the oligonucleotides and
 CC endonuclease are mixed under conditions where an invasive cleavage
 CC structure is formed between the target sequence and the oligonucleotides
 CC if the target sequence is present in the sample, where the invasive
 CC cleavage structure is cleaved by the endonuclease to form a cleavage
 CC product. The composition is useful in detecting and characterising
 CC specific nucleic acid sequences and sequence variants which can be used
 CC in detecting the presence of viral or bacterial infections, and other
 CC diseases such as cancer. The composition may also be used in forensic
 CC analysis or for paternity determinations. The sequence presented is a FEN
 CC -1 related polypeptide used within the scope of the invention.

XX Sequence 527 AA;

Query Match 70.1%; Score 96; DB 5; Length 527;
 Best Local Similarity 87.0%; Pred. No. 1.4e-05;
 Matches 20; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LKQIDAQQTQLRIDSFRLAQOE 23
 |||||:|||||
 Db 345 LKQLNAQQTQLRIDSFRLAQOE 367

RESULT 10
 ADA66104
 ID ADA66104 standard; protein; 527 AA.

XX ADA66104;

XX 20-NOV-2003 (first entry)

XX DNAP-related protein #10.

XX DNAP; invasive cleavage structure; dendrimer; nuclease; endonuclease;
 KW polymerase; polyglycol; 5' nuclease; allelic variation.

XX Xenopus laevis.

XX US2003044796-A1.

XX 06-MAR-2003.
 PD 27-AUG-2001; 2001US-00940244.
 XX 26-NOV-1996; 96US-00756386.
 PR 24-MAR-1998; 98WO-US005809.
 PR 08-JUN-1999; 99US-00350309.
 PR 08-FEB-2000; 2000US-00381212.
 PR 08-DEC-2000; 2000US-00732622.
 XX (NERI/) NERI B P.
 PA (HALL/) HALL J G.
 PA (LYAM/) LYAMICHEV V.
 XX (SMIT/) SMITH L M.
 XX Neri BP, Hall JG, Lyamichev V, Smith LM;
 XX WPI; 2003-596420/56.

XX Detection system for nucleic acid sequences comprises oligonucleotides
 PT configured for hybridizing to target nucleic acid to form invasive
 PT cleavage structure and dendrimer.

XX Disclosure; Fig 59; 354pp; English.

XX The invention relates to a detection system which has oligonucleotides
 CC configured for hybridisation to a target nucleic acid to form an invasive
 CC cleavage structure and dendrimer, where the oligonucleotides are attached
 CC to the dendrimer. The invention also relates to a method for
 CC characterising a nucleic acid sequence comprising providing a sample
 CC suspected of containing a target nucleic acid, oligonucleotides
 CC configured to hybridise to the target nucleic acid to form an invasive
 CC cleavage structure and a dendrimer to which the oligonucleotide is
 CC attached, and exposing the sample to the oligonucleotides and an agent
 CC that detects the presence of an invasive cleavage structure. The agent
 CC comprises a cleavage agent having a structure-specific nuclease.
 CC preferably a 5' nuclease comprising an endonuclease or polymerase. The
 CC detection system further comprises a spacer molecule, consisting of a
 CC carbon chain, polynucleotide or polyglycol, to which the oligonucleotide
 CC is attached. The target molecule and the agent are attached to a solid
 CC support. The invention is used in the detection and characterisation of
 CC nucleic acid sequences and variations in these sequences, used in allelic
 CC variation studies. This sequence represents a protein used in the scope
 CC of the invention.

XX Sequence 527 AA;

Query Match 70.1%; Score 96; DB 7; Length 527;
 Best Local Similarity 87.0%; Pred. No. 1.4e-05;
 Matches 20; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LKQIDAQQTQLRIDSFRLAQOE 23
 |||||:|||||
 Db 345 LKQLNAQQTQLRIDSFRLAQOE 367

RESULT 11
 ABR53625
 ID ABR53625 standard; protein; 493 AA.

XX ABR53625;

XX 20-JUN-2003 (first entry)

XX Protein sequence #SEQ ID 2115.

XX Multiprotein complex; eukaryote; drug target; diagnosis.

XX Saccharomyces cerevisiae.

XX EPI258494-A1.

PD 20-NOV-2002.
 XX
 PF 20-DEC-2001; 2001EP-00130253.
 XX
 PR 15-MAY-2001; 2001EP-00111774.
 XX
 PA (CELL-) CELLZOME AG.
 XX
 PI Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
 PI Marzioch M, Schultz JD, Superti-Furga GD;
 XX
 XX WPI; 2003-250078/25.
 DR N-PSDB; ACC61667.
 XX
 PT New isolated protein complexes useful for diagnosing a disease or
 PT disorder, or as a target for an active agent of a pharmaceutical,
 PT preferably a drug target in the treatment or prevention of disease or
 PT disorder.
 XX
 PS Disclosure; SEQ ID NO 2115; 17pp + Sequence Listing; English.
 XX
 CC The invention relates to multiprotein complexes from eukaryotes. Proteins
 CC of the invention and DNA sequences encoding them are given in records
 CC ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are
 CC obtainable by using a protein as a bait and isolating the set of proteins
 CC which is attached thereto from cells. Such protein complexes may comprise
 CC up to 30 distinct proteins. Protein complexes of the invention are useful
 CC for diagnosing a disease or disorder, or as a target for an active agent
 CC of a pharmaceutical, preferably a drug target in the treatment or
 CC prevention of a disease or disorder. Note: The sequence data for this
 CC patent is not represented in the printed specification, but is based on
 CC sequence information supplied by the European Patent Office. The complete
 CC document is available on CD-ROM
 XX
 XX Sequence 493 AA;

Query Match 41.6%; Score 57; DB 6; Length 493;
 Best Local Similarity 46.2%; Pred. No. 7.7;
 Matches 12; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 2 KQLDAQQTQLRIDSFRLAQOQEKDK 27
 Db 251 KVKDKSQAEIDLFDQFKKAQOQEKDE 276

RESULT 12
 ADK64206
 ID ADK64206 standard; protein; 493 AA.
 XX
 AC ADK64206;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Disease treating protein complex-derived protein #1272.
 XX
 XX protein complex; drug target; diagnosis.
 XX
 OS Unidentified.
 XX
 XX EP1338608-A2.
 XX
 XX 27-AUG-2003.
 XX
 XX 20-DEC-2002; 2002EP-00102902.
 XX
 XX 20-DEC-2001; 2001EP-00130253.
 PR
 XX (CELL-) CELLZOME AG.
 XX
 XX Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;
 PI Marzioch M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;
 PI Michon A, Leutwein C, Rick J;
 XX

DR WPI; 2003-638460/61.
 DR N-PSDB; ADK64207.
 XX
 PT New proteins and protein complexes from eukaryotes, useful as targets in
 PT drug screening, or in diagnosing or screening for the presence of a
 PT disease or disorder, or a predisposition for developing a disease or
 PT disorder in a subject.
 XX
 PS Disclosure; SEQ ID NO 2543; 13pp; English.
 XX
 CC The invention relates to novel protein complexes comprising a first and a
 CC second protein, or its derivative, fragment, homologue or variant. The
 CC proteins are selected from given protein complexes, which are not defined
 CC in the specification. The variants are encoded by nucleic acids that
 CC hybridize to the nucleic acids encoding the proteins under low stringency
 CC conditions. The protein complexes are useful as targets for an active
 CC agent of a pharmaceutical. These protein complexes are particularly
 CC useful as drugs targets for the treatment or preventing of a disease or
 CC disorder. The complexes and methods above are useful in diagnosing or
 CC screening for the presence of a disease or disorder or a predisposition
 CC for developing a disease or disorder in a subject. These are also useful
 CC in screening for a drug for treatment or prevention of a disease or
 CC disorder. The molecule that modulates the amount, activity or protein
 CC components of the complex is useful for the manufacture of a medicament
 CC for the treatment or prevention of a disease or disorder. This sequence
 CC corresponds to a protein of the invention. (Note: the sequence data for
 CC this patent did not form part of the printed specification but was
 CC obtained from the EPO in electronic format).
 XX
 XX Sequence 493 AA;

Query Match 41.6%; Score 57; DB 7; Length 493;
 Best Local Similarity 46.2%; Pred. No. 7.7;
 Matches 12; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 2 KQLDAQQTQLRIDSFRLAQOQEKDK 27
 Db 251 KVKDKSQAEIDLFDQFKKAQOQEKDE 276

RESULT 13
 ADN20626
 ID ADN20626 standard; protein; 1965 AA.
 XX
 AC ADN20626;
 XX
 DT 02-DEC-2004 (first entry)
 XX
 DE Bacterial polypeptide #3279.
 XX
 KW Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polypeptide.
 XX
 OS Bacteria.
 XX
 XX US2003233675-A1.
 XX
 XX 18-DEC-2003.
 PD
 XX 20-FEB-2003; 2003US-00369493.
 XX
 XX 21-FEB-2002; 2002US-0360039P.
 PR
 XX (CAOY/) CAO Y.
 PA (HINK/) HINKLE G J.
 PA (SLAT/) SLATER S C.
 PA (CHEN/) CHEN X.
 PA (GOLD/) GOLDMAN B S.

CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

xx
 SQ Sequence 978 AA;

Query Match 38.0%; Score 52; DB 6; Length 978;
 Best Local Similarity 45.5%; Pred. No. 89;
 Matches 15; Conservative 2; Mismatches 10; Indels 6; Gaps 1;

Qy 1 LKQLDAQQTQLR-----IDSFFRLAQQEKEDK 27
 |||:|||||
 Db 295 LAQLKEATQLPEPKQSDIDAFISLNQGENEIK 327
 |||:|||||

Search completed: May 20, 2005, 10:29:27
 Job time : 164 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2005, 10:25:14 ; Search time 43 Seconds
(without alignments)
48.609 Million cell updates/sec

Title: US-10-083-576-1

Perfect score: 137

Sequence: 1 LKQLDAQOTQLRIDSFFRLAQEKEDKR 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pap.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	128	93.4	550	2	US-08-823-516-142
2	128	93.4	550	4	US-09-940-244-142
3	110	80.3	543	2	US-08-823-516-143
4	110	80.3	543	4	US-09-940-244-143
5	96	70.1	527	2	US-08-823-516-144
6	96	70.1	527	4	US-09-940-244-144
7	56	40.9	422	4	US-09-248-796A-19321
8	53.5	39.1	121	4	US-09-270-767-56735
9	53.5	39.1	336	4	US-09-270-767-41511
10	50	36.5	409	4	US-09-902-540-12745
11	49	35.8	288	4	US-09-252-991A-29956
12	48	35.0	439	4	US-09-248-796A-22837
13	47	34.3	478	4	US-09-543-681A-5186
14	46.5	33.9	351	1	US-08-402-217A-2
15	46.5	33.9	351	1	US-08-700-178-2
16	46.5	33.9	351	3	US-08-995-654-2
17	46	33.6	556	3	US-08-687-590-24
18	46	33.6	563	4	US-09-949-016-10046
19	46	33.6	563	4	US-09-949-016-10047
20	45	32.8	133	4	US-09-513-999C-7956
21	45	32.8	191	3	US-08-655-352-7
22	45	32.8	191	3	US-09-258-016-7
23	45	32.8	191	4	US-09-257-825B-7
24	45	32.8	1156	4	US-09-107-532A-6394
25	45	32.8	1156	4	US-09-198-452A-171
26	45	32.8	1562	4	US-09-438-185A-152
27	44.5	32.5	134	4	US-09-270-767-60370

28	44.5	32.5	263	4	US-09-902-540-11627	Sequence 11627, A
29	44.5	32.5	601	4	US-09-252-991A-27821	Sequence 27821, A
30	44.5	32.5	676	4	US-09-270-767-44900	Sequence 44900, A
31	44	32.1	130	4	US-09-621-976-5203	Sequence 5203, Ap
32	44	32.1	141	4	US-09-621-976-3895	Sequence 3895, Ap
33	44	32.1	335	4	US-09-248-796A-22784	Sequence 22784, A
34	44	32.1	368	4	US-09-328-352-7910	Sequence 7910, Ap
35	44	32.1	435	4	US-09-198-452A-603	Sequence 603, App
36	44	32.1	1405	4	US-09-438-185A-566	Sequence 566, App
37	43	31.4	189	4	US-09-082-358B-8	Sequence 8, Appl
38	43	31.4	236	4	US-09-640-211A-1063	Sequence 1063, Ap
39	43	31.4	257	2	US-08-867-087B-2	Sequence 2, Appl
40	43	31.4	341	4	US-09-902-540-11473	Sequence 11473, A
41	43	31.4	352	4	US-09-543-681A-4723	Sequence 4723, Ap
42	43	31.4	369	4	US-09-489-039A-13829	Sequence 13829, A
43	43	31.4	401	4	US-09-252-991A-17272	Sequence 17272, A
44	43	31.4	613	4	US-09-543-681A-6012	Sequence 6012, Ap
45	43	31.4	658	4	US-09-248-796A-20278	Sequence 20278, A

ALIGNMENTS

RESULT 1
US-08-823-516-142
; Sequence 142, Application US/08823516
; Patent No. 5994069
; GENERAL INFORMATION:
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Mast, Andrea L.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple
; TITLE OF INVENTION: Sequential Invasive Cleavages
; NUMBER OF SEQUENCES: 163
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/823,516
; FILING DATE: 24-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/01072
; FILING DATE: 21-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/759,038
; FILING DATE: 02-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/758,314
; FILING DATE: 02-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/756,386
; FILING DATE: 29-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/682,853
; FILING DATE: 12-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,491
; FILING DATE: 24-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-02736

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 550 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-08-823-516-142

Query Match          93.4%; Score 128; DB 2; Length 550;
Best Local Similarity 92.9%; Pred. No. 3e-11;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LKQIDAQQTQLRIDSPFRLAQKEKEDKR 28
Db 345 LKQIDAQQTQLRIDSPFRLAQKEKEDAK 372

RESULT 2
US-09-940-244-142
; Sequence 142, Application US/09940244
; Patent No. 6692917
; GENERAL INFORMATION:
; APPLICANT: Neri, Bruce P.
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor
; APPLICANT: Smith, Lloyd M.
; TITLE OF INVENTION: Reactions on Dendrimers
; FILE REFERENCE: FORS-06478
; CURRENT APPLICATION NUMBER: US/09/940,244
; CURRENT FILING DATE: 2002-05-06
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 142:
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-940-244-142

Query Match          93.4%; Score 128; DB 4; Length 550;
Best Local Similarity 92.9%; Pred. No. 3e-11;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LKQIDAQQTQLRIDSPFRLAQKEKEDKR 28
Db 345 LKQIDAQQTQLRIDSPFRLAQKEKEDAK 372

RESULT 3
US-08-823-516-143
; Sequence 143, Application US/08823516
; Patent No. 5994069
; GENERAL INFORMATION:
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Mast, Andrea L.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: Detection of Nucleic Acids By Multiple
; TITLE OF INVENTION: Sequential Invasive Cleavages
; NUMBER OF SEQUENCES: 163
; CORRESPONDENCE ADDRESS:
; ADDRESS: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/823,516
; FILING DATE: 24-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/01072
; FILING DATE: 21-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/759,038
; FILING DATE: 02-DEC-1996
; APPLICATION NUMBER: US 08/758,314
; FILING DATE: 02-DEC-1996
; APPLICATION NUMBER: US 08/756,386
; FILING DATE: 29-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/682,853
; FILING DATE: 12-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,491
; FILING DATE: 24-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-02736
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-08-823-516-143

Query Match          80.3%; Score 110; DB 2; Length 543;
Best Local Similarity 78.6%; Pred. No. 1.6e-08;
Matches 22; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LKQIDAQQTQLRIDSPFRLAQKEKEDKR 28
Db 345 LKELNAHQTLRIDSPFRLAQKEKQDAK 372

RESULT 4
US-09-940-244-143
; Sequence 143, Application US/09940244
; Patent No. 6692917
; GENERAL INFORMATION:
; APPLICANT: Neri, Bruce P.
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor
; APPLICANT: Smith, Lloyd M.
; TITLE OF INVENTION: Reactions on Dendrimers
; FILE REFERENCE: FORS-06478
; CURRENT APPLICATION NUMBER: US/09/940,244
; CURRENT FILING DATE: 2002-05-06
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 143:
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-940-244-143

Query Match          80.3%; Score 110; DB 4; Length 543;
Best Local Similarity 78.6%; Pred. No. 1.6e-08;
Matches 22; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
```

QY 1 LKQLDAQQTQLRIDSFFRLAQOEKEDK 28
 Db 345 LKGLNAHOTQLRIDSFFRLAQOEKQDAK 372

RESULT 5
 US-08-823-516-144
 ; Sequence 144, Application US/08823516
 ; Patent No. 5994069
 ; GENERAL INFORMATION:
 ; APPLICANT: Hall, Jeff G.
 ; APPLICANT: Lyamichev, Victor I.
 ; APPLICANT: Mast, Andrea L.
 ; APPLICANT: Brow, Mary Ann D.
 ; TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple
 ; TITLE OF INVENTION: Sequential Invasive Cleavages
 ; NUMBER OF SEQUENCES: 163
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Medlen & Carroll, LLP
 ; STREET: 220 Montgomery Street, Suite 2200
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States Of America
 ; ZIP: 94104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/823,516
 ; FILING DATE: 24-MAR-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US97/01072
 ; FILING DATE: 21-JAN-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/759,038
 ; FILING DATE: 02-DEC-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/758,314
 ; FILING DATE: 02-DEC-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/756,386
 ; FILING DATE: 29-NOV-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/682,853
 ; FILING DATE: 12-JUL-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/599,491
 ; FILING DATE: 24-JAN-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ingolia, Diane E.
 ; REGISTRATION NUMBER: 40,027
 ; REFERENCE/DOCKET NUMBER: FORS-02736
 ; TELEPHONE: (415) 705-8410
 ; TELEFAX: (415) 397-8338
 ; INFORMATION FOR SEQ ID NO: 144:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 527 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: not relevant
 ; TOPOLOGY: not relevant
 ; MOLECULE TYPE: protein
 ; US-08-823-516-144

Query Match 70.1%; Score 96; DB 2; Length 527;
 Best Local Similarity 87.0%; Pred. No. 2e-06;
 Matches 20; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LKQLDAQQTQLRIDSFFRLAQOE 23

Db 345 LKQLNAQQTQLRIDSFFRLAQOE 367

RESULT 6
 US-09-940-244-144
 ; Sequence 144, Application US/09940244
 ; Patent No. 6692917
 ; GENERAL INFORMATION:
 ; APPLICANT: Neri, Bruce P.
 ; APPLICANT: Hall, Jeff G.
 ; APPLICANT: Lyamichev, Victor
 ; APPLICANT: Smith, Lloyd M.
 ; TITLE OF INVENTION: Reactions on Dendrimers
 ; FILE REFERENCE: FORS-06478
 ; CURRENT APPLICATION NUMBER: US/09/940,244
 ; CURRENT FILING DATE: 2002-05-06
 ; NUMBER OF SEQ ID NOS: 422
 ; SOFTWARE: Patent In version 3.1
 ; SEQ ID NO 144
 ; LENGTH: 527
 ; TYPE: PRT
 ; ORGANISM: Xenopus laevis
 ; US-09-940-244-144

Query Match 70.1%; Score 96; DB 4; Length 527;
 Best Local Similarity 87.0%; Pred. No. 2e-06;
 Matches 20; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LKQLDAQQTQLRIDSFFRLAQOE 23
 Db 345 LKQLNAQQTQLRIDSFFRLAQOE 367

RESULT 7
 US-09-248-796A-19321
 ; Sequence 19321, Application US/09248796A
 ; Patent No. 6747137
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith Weinstock et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
 ; FILE REFERENCE: 107196.132
 ; CURRENT APPLICATION NUMBER: US/09/248,796A
 ; CURRENT FILING DATE: 1999-02-12
 ; PRIOR APPLICATION NUMBER: US 60/074,725
 ; PRIOR FILING DATE: 1998-02-13
 ; PRIOR APPLICATION NUMBER: US 60/096,409
 ; PRIOR FILING DATE: 1998-08-13
 ; NUMBER OF SEQ ID NOS: 28208
 ; SEQ ID NO 19321
 ; LENGTH: 422
 ; TYPE: PRT
 ; ORGANISM: Candida albicans
 ; US-09-248-796A-19321

Query Match 40.9%; Score 56; DB 4; Length 422;
 Best Local Similarity 42.9%; Pred. No. 1.7;
 Matches 9; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 7 QQTQLRIDSFFRLAQOEKEDK 27
 Db 219 ERSQMKISSFFQIGQKREKK 239

RESULT 8
 US-09-270-767-56735
 ; Sequence 56735, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 56735
; LENGTH: 121
; TYPE: PR1
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-56735

Query Match 39.1%; Score 53.5; DB 4; Length 121;
Best Local Similarity 44.0%; Pred. No. 0.96;
Matches 11; Conservative 7; Mismatches 6; Indels 1; Gaps 1;

QY 5 DAQQTQLRIDSF-FRLAQKEKDKR 28
DB 16 EKEQAQRVDYRRRLTEQEHDKR 40

RESULT 9
US-09-270-767-41511
; Sequence 41511, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41511
; LENGTH: 336
; TYPE: PR1
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-41511

Query Match 39.1%; Score 53.5; DB 4; Length 336;
Best Local Similarity 44.0%; Pred. No. 3.1;
Matches 11; Conservative 7; Mismatches 6; Indels 1; Gaps 1;

QY 5 DAQQTQLRIDSF-FRLAQKEKDKR 28
DB 231 EKEQAQRVDYRRRLTEQEHDKR 255

RESULT 10
US-09-902-540-12745
; Sequence 12745, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 12745
; LENGTH: 409
; TYPE: PR1
; ORGANISM: Myxococcus xanthus
US-09-902-540-12745

Query Match 36.5%; Score 50; DB 4; Length 409;

Best Local Similarity 38.5%; Pred. No. 13;
Matches 10; Conservative 7; Mismatches 9; Indels 0; Gaps 0;
QY 3 QLDAAQTQLRIDSF-FRLAQKEKDKR 28
DB 163 RMDAQRQHEAEQLRLAQQAESR 188

RESULT 11
US-09-252-991A-29966
; Sequence 29966, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29966
; LENGTH: 288
; TYPE: PR1
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29966

Query Match 35.8%; Score 49; DB 4; Length 288;
Best Local Similarity 43.3%; Pred. No. 13;
Matches 13; Conservative 5; Mismatches 10; Indels 2; Gaps 1;

QY 1 LKQDAQQTQLRIDSF--RLAQKEKDKR 28
DB 158 LVELETQSRQLRVKSEFWSRVALEPEQAR 187

RESULT 12
US-09-248-796A-22837
; Sequence 22837, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN.
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 22837
; LENGTH: 439
; TYPE: PR1
; ORGANISM: Candida albicans
US-09-248-796A-22837

Query Match 35.0%; Score 48; DB 4; Length 439;
Best Local Similarity 40.7%; Pred. No. 29;
Matches 11; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 1 LKQDAQQTQLRIDSF-FRLAQKEKDK 27
DB 315 LKSLDVRCNLRVISPFLVSMLESENK 341

RESULT 13
US-09-543-681A-5186
; Sequence 5186, Application US/09543681A

Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
PRIOR FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 5186
LENGTH: 478
TYPE: PR1
ORGANISM: Proteus mirabilis
US-09-543-681A-5186

Query Match 34.3%; Score 47; DB 4; Length 478;
Best Local Similarity 36.4%; Pred. No. 45;
Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 6 AOOTQLRIDSPFRLAQKEK 27
DB 271 AOOTQINSKFKYQKLEHEWDR 292

RESULT 14
US-08-402-217A-2
Sequence 2, Application US/08402217A
Patent No. 5587301
GENERAL INFORMATION:
APPLICANT: Hawkins, Phillip R.
APPLICANT: Wilde, Craig G.
APPLICANT: Seilhamer, Jeffrey J.
TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/402,217A
FILING DATE: 10-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Luther, Barbara J.
REGISTRATION NUMBER: 33954
REFERENCE/DOCKET NUMBER: PF-0028US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-852-0195
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 351 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-402-217A-2

Query Match 33.9%; Score 46.5; DB 1; Length 351;
Best Local Similarity 63.2%; Pred. No. 38;
Matches 12; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

QY 7 QOTQLRIDSPFRLAQKE 25

Db 19 QQKELQIDS---LLQKE 34

RESULT 15
US-08-700-178-2
Sequence 2, Application US/08700178
Patent No. 5783669
Patent No. 5783669 5700912
GENERAL INFORMATION:
APPLICANT: Hawkins, Phillip R.
APPLICANT: Wilde, Craig G.
APPLICANT: Seilhamer, Jeffrey J.
TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/700,178
FILING DATE: August 20, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/402,217
FILING DATE: March 10, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0028-1 DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 351 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-700-178-2

Query Match 33.9%; Score 46.5; DB 1; Length 351;
Best Local Similarity 63.2%; Pred. No. 38;
Matches 12; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

QY 7 QOTQLRIDSPFRLAQKE 25

Db 19 QQKELQIDS---LLQKE 34

Search completed: May 20, 2005, 10:30:15
Job time : 44 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2005, 10:25:10 ; Search time 16 Seconds
(without alignments)
168.379 Million cell updates/sec

Title: US-10-083-576-1

Perfect score: 137

Sequence: 1 LKQLDAAQQTOLRIDSPFRLAQOEKDKR 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_79.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	128	93.4	1185	2 I58009	Gene ERCC5 protein
2	128	93.4	1186	2 S35993	DNA repair protein
3	110	80.3	1170	2 A57650	repair protein XPG
4	96	70.1	1196	2 S35994	DNA repair protein
5	60	43.8	492	2 F86263	hypothetical prote
6	57	41.6	493	2 S50625	GL03 protein - yea
7	53	38.7	129	2 T41014	probable proteasom
8	52	38.0	978	2 B89971	conserved hypotet
9	51	37.2	336	2 S61299	lipopolysaccharide
10	50	36.5	355	2 S44507	hypothetical prote
11	49	35.8	276	2 B83430	hypothetical prote
12	49	35.8	355	2 T40597	hypothetical prote
13	49	35.8	1451	2 B86286	F9L1.15 protein -
14	48	35.0	189	2 A81282	probable periplasm
15	48	35.0	229	2 T22037	hypothetical prote
16	47	34.3	238	2 AG2503	hypothetical prote
17	47	34.3	385	2 T13692	hypothetical prote
18	47	34.3	673	2 T36717	probable serine/th
19	47	34.3	1400	2 T31555	hypothetical prote
20	47	34.3	1469	2 H96622	probable ABC trans
21	46.5	33.9	725	1 JC5016	hyaluronan recepto
22	46	33.6	39	2 A48911	gene Tcp-ix protei
23	46	33.6	345	2 T16935	hypothetical prote
24	46	33.6	384	2 A81758	integrases homolog
25	46	33.6	556	2 S10486	t-complex-type mol
26	46	33.6	556	2 JC1443	t-complex polypept
27	46	33.6	556	2 JC0866	T-complex protein
28	46	33.6	1102	2 AB1940	methyl-accepting c
29	46	33.6	1108	2 AF1047	probable membrane

30	46	33.6	1274	2 T39249	probable tripeptid
31	45.5	33.2	374	2 C84040	hypothetical prote
32	45.5	33.2	414	2 T26205	hypothetical prote
33	45.5	33.2	835	1 I57441	involucrin - orang
34	45.5	33.2	963	2 H69818	conserved hypotet
35	45.5	33.2	1087	2 T49496	hypothetical prote
36	45	32.8	165	2 A44103	neurocalcin beta -
37	45	32.8	191	2 JH0815	neural visinin-lik
38	45	32.8	336	2 C81073	ADP-heptose-LPS he
39	45	32.8	336	2 B81869	lipopolysaccharide
40	45	32.8	406	2 A35360	otefin - fruit fly
41	45	32.8	556	2 S13163	t-complex-type mol
42	45	32.8	1537	2 F86509	CR147 hypothetical
43	45	32.8	1537	2 C81558	conserved hypotet
44	45	32.8	1537	2 H72112	ct147 hypothetical
45	44.5	32.5	167	2 T00298	hypothetical prote

ALIGNMENTS

RESULT 1

I58009

Gene ERCC5 protein - human

C/Species: Homo sapiens (man)

C/Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000

C/Accession: I58009

R/Shiomi, T.; Harada, Y.

Mutat. Res. 314, 167-175, 1994

A/Title: An ERCC5 gene with homology to yeast RAD 2 is involved in group G xeroderma pig

A/Reference number: I58009; MUID:94173288; PMID:7510366

A/Accession: I58009

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-1185 <RES>

A/Cross-references: GB:D16305; NID:G303606; PIDN:BAA03812.1; PID:G303607

C/Genetics:

A/Gene: GDB:ERCC5

A/Cross-references: GDB:120515; OMIM:133530

A/Map position: 13q33-13q33

Query Match 93.4%; Score 128; DB 2; Length 1185;
Best Local Similarity 92.9%; Pred. No. 2.4e-10;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LKQLDAAQQTOLRIDSPFRLAQOEKDKR 28
Db 980 LKQLDAAQQTOLRIDSPFRLAQOEKEDAK 1007

RESULT 2

S35993

DNA repair protein XPG - human

N/Alternate names: xeroderma pigmentosum group G complementing factor; XP-G factor

C/Species: Homo sapiens (man)

C/Date: 22-Nov-1993 #sequence_revision 12-Apr-1996 #text_change 09-Jul-2004

C/Accession: S35993; S33319; A54439; I39304

R/Clarkson, S.G.

submitted to the EMBL Data Library, January 1993

A/Reference number: S35993

A/Accession: S35993

A/Molecule type: mRNA

A/Residues: 1-1186 <CLA>

A/Cross-references: UNIPROT:P28715; EMBL:X69978; NID:G298110; PID:G298111

R/Scherly, D.; Nusspikel, T.; Corlet, J.; Ucla, C.; Bairoch, A.; Clarkson, S.G.

Nature 363, 182-185, 1993

A/Title: Complementatation of the DNA repair defect in xeroderma pigmentosum group G cells

A/Reference number: S33319; MUID:93247645; PMID:8483504

A/Accession: S33319

A/Status: nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 1-95,753-889,1166-1186 <SCH>

A/Cross-references: EMBL:X69978

R;MacInnes, M.A.; Dickson, J.A.; Hernandez, R.R.; Learmonth, D.; Lin, G.Y.; Mudgett, J.S.
Mol. Cell. Biol. 13, 6393-6402, 1993
A:Title: Human ERCC5 cDNA-cosmid complementation for excision repair and bipartite amino
A:Reference number: A54439; MUID:94019313; PMID:8413238
A:Accession: A54439
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-253, 'V', 255-1103, 'D', 1105-1186 <MAC>
A:Cross-references: GB:L20046; NID:9306741; PIDN:AAC37533.1; PID:9306742
A>Note: neither nucleotide sequence nor conceptual translation is complete
R;Samec, S.; Jones, T.A.; Corlet, J.; Scherly, D.; Sheer, D.; Wood, R.D.; Clarkson, S.G.
Genomics 21, 283-285, 1994
A:Title: The human gene for xeroderma pigmentosum complementation group G (XPG) maps to
A:Reference number: I39304; MUID:94375034; PMID:8088806
A:Accession: I39304
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-88 <RES>
A:Cross-references: EMBL:X71341; NID:9495253; PIDN:CAA50481.1; PID:9825732
C:Genetics:
A:Gene: GDB:ERCC5; ERCC2
A:Cross-references: GDB:120515; OMIM:133530
A:Map position: 13q33-13q33
A:Introns: 30/1
A>Note: the list of introns is incomplete
C:Keywords: DNA binding

Query Match 93.4%; Score 128; DB 2; Length 1186;
Best Local Similarity 92.9%; Pred. No. 2.4e-10;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LKQDAQQTQLRIDSFRLAQOEKEDKR 28
DB 981 LKQDAQQTQLRIDSFRLAQOEKEDAK 1008

RESULT 3
A57650
repair protein XPG - mouse
C:Species: Mus musculus (house mouse)
C>Date: 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change 09-Jul-2004
A:Reference number: A57650
R;Harada, Y.N.; Matsuda, Y.; Shiomi, N.; Shiomi, T.
Genomics 28, 59-65, 1995
A:Title: Complementary DNA sequence and chromosomal localization of xpg, the mouse count
A:Reference number: A57650; MUID:96070433; PMID:7590748
A:Accession: A57650
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1170 <HAF>
A:Cross-references: UNIPROT:P35689; GB:D16306; NID:91389576; PIDN:BAA03813.1; PID:d10043
C:Genetics:
A:Gene: xpg
C:Keywords: DNA repair

Query Match 80.3%; Score 110; DB 2; Length 1170;
Best Local Similarity 78.6%; Pred. No. 9.5e-08;
Matches 22; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LKQDAQQTQLRIDSFRLAQOEKEDKR 28
DB 980 LKHLNAHQTLRIDSFRLAQOEKQDAK 1007

RESULT 4
S35994
DNA repair protein XPGC - African clawed frog
N:Alternate names: XP-G-related factor
C:Species: Xenopus laevis (African clawed frog)
C>Date: 03-Mar-1994 #sequence_revision 12-Apr-1996 #text_change 09-Jul-2004
A:Reference number: S35994; S33320
R;Clarkson, S.G.
submitted to the EMBL Data Library, January 1993

A:Reference number: S35993
A:Accession: S35994
A:Molecule type: mRNA
A:Residues: 1-1196 <CLA>
A:Cross-references: UNIPROT:P14629; EMBL:X69977; NID:9312432; PID:9312433
R;Scherly, D.; Nusspikler, T.; Corlet, J.; Ucia, C.; Bairoch, A.; Clarkson, S.G.
Nature 363, 182-185, 1993
A:Title: Complementation of the DNA repair defect in xeroderma pigmentosum group G cells
A:Reference number: S33319; MUID:93247645; PMID:8483504
A:Accession: S33320
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-95, 786-922, 1176-1196 <SCH>
A:Cross-references: EMBL:X69977
C:Keywords: DNA binding

Query Match 70.1%; Score 96; DB 2; Length 1196;
Best Local Similarity 87.0%; Pred. No. 1e-05;
Matches 20; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LKQDAQQTQLRIDSFRLAQOE 23
DB 1014 LKQDAQQTQLRIDSFRLQOHE 1036

RESULT 5
F86263
hypothetical protein F13K23.23 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
A:Accession: F86263
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: F86263
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-492 <STO>
A:Cross-references: UNIPROT:Q9LPV2; GB:AE005172; NID:98698744; PIDN:AAF78502.1; GSPDB:GN
C:Genetics:
A:Map position: 1

Query Match 43.8%; Score 60; DB 2; Length 492;
Best Local Similarity 50.0%; Pred. No. 0.66;
Matches 13; Conservative 5; Mismatches 6; Indels 2; Gaps 1;

QY 1 LKQDAQQTQLRI--DSFFRLAQOEK 24
DB 378 LRELDLNNQIRVLPDSFFRLKLEK 403

RESULT 6
S50625
GLO3 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YER122c
C:Species: Saccharomyces cerevisiae
C>Date: 28-Jan-1995 #sequence_revision 12-May-1995 #text_change 09-Jul-2004
A:Reference number: S50625; S47007; S50036
R;Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmid 9781, 8198, 9115, 9981, and lambda
A:Reference number: S50625
A:Accession: S50625
A:Molecule type: DNA
A:Residues: 1-493 <DIE>

lipopolysaccharide heptosyltransferase (EC 2.4.99.-) II rfaF [similarity] - Neisseria gonorrhoeae

QY 1 LKQLDAQOQTOLRIDSFRRLAQOEKED 26

C;Species: Bacteroides fragilis
C:Date: 16-Feb-1995 #sequence revision 09-Mar-1996 #text change 09-Jul-2004

A;Residues: 1-355 <POD>
 A;Cross-references: UNIPROT: Q09536, EMBL: Y722A1, NID: Q312686, PDB: 1A2E10A9, 1. PDB: Q312686

[illegible]

```
A;Reference number: 249303
A;Accession: I22037
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-229 <WIL>
A;Cross-references: UNIPROT:Q20231; EMBL:Z69302; PIDN:CAA93265.1; GSPDB:GN00020; CESP:F4(
A;Experimental source: clone F40F8
C;Genetics:
A;Gene: CESP:F40F8.3
A;Map position: 2
A;Introns: 25/1; 61/3; 105/3; 126/1
C;Superfamily: Caenorhabditis elegans hypothetical protein F40F8.3

Query Match      35,0%; Score 48; DB 2; Length 229;
Best Local Similarity    29,6%; Pred. No. 16;
Matches      8; Conservative   10; Mismatches     9; Indels     0; Gaps     0;

Qy      1 LKQLDAQQTQLRIDSPFFRLAQOKEEDK 27
        ||| : |||::|| ::||
Db       4 IKKLSSKKYFSLLKLNFKVRKENMKD 30
```

Search completed: May 20, 2005, 10:26:38
Job time : 17 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2005, 10:26:45 ; Search time 1925 Seconds
(without alignments)
4.866 Million cell updates/sec

Title: US-10-083-576-1

Perfect score: 137

Sequence: 1 LKQLDAQOQTQLRIDSFFRLAQOEKEDKR 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1434725 seqs, 334507595 residues

Total number of hits satisfying chosen parameters: 1434725

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	128	93.4	550	10	US-09-940-244-142
2	128	93.4	550	13	US-10-033-297-144
3	128	93.4	550	14	US-10-290-386-142
4	128	93.4	550	15	US-10-356-861-142
5	128	93.4	550	16	US-10-897-793-142
6	128	93.4	550	17	US-10-783-557-143
7	128	93.4	1186	15	US-10-341-434-51
8	110	80.3	543	10	US-09-940-244-143
9	110	80.3	543	13	US-10-033-297-143
10	110	80.3	543	14	US-10-290-386-143
11	110	80.3	543	15	US-10-356-861-143
12	110	80.3	543	16	US-10-897-793-143
13	110	80.3	543	17	US-10-783-557-144

14	96	70.1	527	10	US-09-940-244-144	Sequence 144, App
15	96	70.1	527	13	US-10-033-297-144	Sequence 144, App
16	96	70.1	527	14	US-10-290-386-144	Sequence 144, App
17	96	70.1	527	15	US-10-356-861-144	Sequence 144, App
18	96	70.1	527	16	US-10-897-793-144	Sequence 144, App
19	96	70.1	527	17	US-10-783-557-145	Sequence 145, App
20	56	40.9	1965	15	US-10-369-493-3279	Sequence 3279, App
21	52	38.0	978	15	US-10-282-122A-44151	Sequence 44151, A
22	52	38.0	981	9	US-09-815-242-12211	Sequence 12211, A
23	51	37.2	132	16	US-10-767-701-58982	Sequence 58982, A
24	50.5	36.9	165	15	US-10-424-599-196749	Sequence 196749, A
25	50	36.5	55	16	US-10-437-963-166573	Sequence 166573, A
26	50	36.5	246	16	US-10-767-701-46498	Sequence 46498, A
27	50	36.5	1457	15	US-10-424-599-227147	Sequence 227147, A
28	48	35.0	225	16	US-10-767-701-47163	Sequence 47163, A
29	48	35.0	294	15	US-10-425-114-52116	Sequence 52116, A
30	48	35.0	412	15	US-10-282-122A-61900	Sequence 61900, A
31	47	34.3	518	9	US-09-925-299-862	Sequence 862, App
32	47	34.3	518	10	US-09-925-299-862	Sequence 862, App
33	47	34.3	532	15	US-10-424-599-172364	Sequence 172364, App
34	47	34.3	555	14	US-10-106-698-5847	Sequence 5847, App
35	47	34.3	582	16	US-10-437-963-178712	Sequence 178712, App
36	47	34.3	667	10	US-09-863-776-16	Sequence 16, Appl
37	47	34.3	667	10	US-09-863-776-18	Sequence 18, Appl
38	47	34.3	670	10	US-09-823-187-86	Sequence 86, Appl
39	47	34.3	670	10	US-09-863-776-51	Sequence 51, Appl
40	47	34.3	671	10	US-09-946-374-308	Sequence 308, App
41	47	34.3	671	10	US-09-823-187-85	Sequence 85, Appl
42	47	34.3	671	10	US-09-863-776-20	Sequence 20, Appl
43	47	34.3	671	13	US-10-052-586-346	Sequence 346, App
44	47	34.3	671	14	US-10-174-590-346	Sequence 346, App
45	47	34.3	671	14	US-10-176-758-346	Sequence 346, App

ALIGNMENTS

RESULT 1

US-09-940-244-142
; Sequence 142, Application US/09940244
; Publication No. US2003004796A1
; GENERAL INFORMATION:
; APPLICANT: Neri, Bruce P.
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor
; APPLICANT: Smith, Lloyd M.
; TITLE OF INVENTION: Reactions on Dendrimers
; FILE REFERENCE: FORS-06478
; CURRENT APPLICATION NUMBER: US/09/940,244
; CURRENT FILING DATE: 2002-05-06
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 142
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-940-244-142

Query Match 93.4%; Score 128; DB 10; Length 550;

Best Local Similarity 92.9%; Pred. No. 2.8e-10;

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LKQLDAQOQTQLRIDSFFRLAQOEKEDKR 28

Db 345 LKQLDAQOQTQLRIDSFFRLAQOEKEDAK 372

RESULT 2

US-10-033-297-142
; Sequence 142, Application US/10033297
; Publication No. US20020187486A1
; GENERAL INFORMATION:
; APPLICANT: Hall, Jeff G.



Db 345 LKQLDAQQTQLRIDSFFRLAQOQEKEDAK 372

RESULT 5

US-10-897-793-142

Sequence 142, Application US/10897793

Publication No. US20050003432A1

GENERAL INFORMATION:

APPLICANT: Hall, Jeff G.

APPLICANT: Lyamichiev, Victor I.

Maest, Andrea L.

Brow, Mary Ann D.

TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple Sequential Invasive Cleavages

NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medlen & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States Of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/897,793

FILING DATE: 23-Jul-2004

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US97/01072

FILING DATE: 21-JAN-1997

APPLICATION NUMBER: US 08/759,038

FILING DATE: 02-DEC-1996

APPLICATION NUMBER: US 08/758,314

FILING DATE: 02-DEC-1996

APPLICATION NUMBER: US 08/756,386

FILING DATE: 29-NOV-1996

APPLICATION NUMBER: US 08/682,853

FILING DATE: 12-JUL-1996

APPLICATION NUMBER: US 08/599,491

FILING DATE: 24-JAN-1996

APPLICATION NUMBER: US 08/823,516

FILING DATE: 24-MAR-1997

ATTORNEY/AGENT INFORMATION:

NAME: MacKnight, Kamrin T.

REGISTRATION NUMBER: 38,230

REFERENCE/DOCKET NUMBER: FORS-03295

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 142:

SEQUENCE CHARACTERISTICS:

LENGTH: 550 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 142:

US-10-897-793-142

Query Match 93.4%; Score 128; DB 16; Length 550;

Best Local Similarity 92.9%; Pred. No. 2.8e-10;

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LKQLDAQQTQLRIDSFFRLAQOQEKEDKR 28

Db 345 LKQLDAQQTQLRIDSFFRLAQOQEKEDAK 372

RESULT 6

US-10-783-557-143

Sequence 143, Application US/10783557

Publication No. US20050048527A1

GENERAL INFORMATION:

APPLICANT: Allawi, Hatim T.

APPLICANT: Kaiser, Michael W.

APPLICANT: Ma, Wu-Po

APPLICANT: Neri, Bruce P.

APPLICANT: Lyamichiev, Victor I.

TITLE OF INVENTION: Endonuclease-Substrate Complexes

FILE REFERENCE: FORS-08907

CURRENT APPLICATION NUMBER: US/10/783,557

CURRENT FILING DATE: 2004-02-20

NUMBER OF SEQ ID NOS: 533

SOFTWARE: Patent In version 3.2

SEQ ID NO 143

LENGTH: 550

TYPE: PRT

ORGANISM: Homo sapiens

US-10-783-557-143

Query Match 93.4%; Score 128; DB 17; Length 550;

Best Local Similarity 92.9%; Pred. No. 2.8e-10;

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LKQLDAQQTQLRIDSFFRLAQOQEKEDKR 28

Db 345 LKQLDAQQTQLRIDSFFRLAQOQEKEDAK 372

RESULT 7

US-10-341-434-51

Sequence 51, Application US/10341434

Publication No. US20030215835A1

GENERAL INFORMATION:

APPLICANT: Origene Technologies

TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes

FILE REFERENCE: 9U 204 205 R1

CURRENT APPLICATION NUMBER: US/10/341,434

CURRENT FILING DATE: 2003-07-18

PRIOR APPLICATION NUMBER: US 60/348,164

PRIOR FILING DATE: 2002-01-15

PRIOR APPLICATION NUMBER: US 60/348,119

PRIOR FILING DATE: 2002-01-15

NUMBER OF SEQ ID NOS: 238

SOFTWARE: Patent In version 3.1

SEQ ID NO 51

LENGTH: 1186

TYPE: PRT

ORGANISM: Homo sapiens

US-10-341-434-51

Query Match 93.4%; Score 128; DB 15; Length 1186;

Best Local Similarity 92.9%; Pred. No. 6.8e-10;

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LKQLDAQQTQLRIDSFFRLAQOQEKEDKR 28

Db 981 LKQLDAQQTQLRIDSFFRLAQOQEKEDAK 1008

RESULT 8

US-09-940-244-143

Sequence 143, Application US/09940244

Publication No. US20030044796A1

GENERAL INFORMATION:

APPLICANT: Neri, Bruce P.

APPLICANT: Hall, Jeff G.

APPLICANT: Lyamichiev, Victor

APPLICANT: Smith, Lloyd M.

TITLE OF INVENTION: Reactions on Dendrimers

FILE REFERENCE: FORS-06478

CURRENT APPLICATION NUMBER: US/09/940,244

```
; CURRENT FILING DATE: 2002-05-06
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 143
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-940-244-143

Query Match      80.3%; Score 110; DB 10; Length 543;
Best Local Similarity 78.6%; Pred. No. 1.2e-07;
Matches 22; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LKQLDAQQTQLRIDSFFRLAQKEKDKR 28
   |||:|||||:|||||:|||||:|||||:
Db 345 LKHLNAHQTLRIDSFFRLAQKEKQDAK 372

RESULT 9
US-10-033-297-143
; Sequence 143, Application US/10033297
; Publication No. US20020187486A1
; GENERAL INFORMATION:
; APPLICANT: Hall, Jeff G.
; LYAMICHEV, Victor I.
; Mast, Andrea L.
; Brow, Mary Ann D.
; TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple
; Sequential Invasive Cleavages
; NUMBER OF SEQUENCES: 163
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/033,297
; FILING DATE: 12-NOV-1996
; CLASSIFICATION: <unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/09/350,597
; FILING DATE: 09-JUL-1999
; APPLICATION NUMBER: US/08/823,516
; FILING DATE: 24-MAR-1997
; APPLICATION NUMBER: PCT/US97/01072
; FILING DATE: 21-JAN-1997
; APPLICATION NUMBER: US 08/759,038
; FILING DATE: 02-DEC-1996
; APPLICATION NUMBER: US 08/758,314
; FILING DATE: 02-DEC-1996
; APPLICATION NUMBER: US 08/756,386
; FILING DATE: 29-NOV-1996
; APPLICATION NUMBER: US 08/682,853
; FILING DATE: 12-JUL-1996
; APPLICATION NUMBER: US 08/599,491
; FILING DATE: 24-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-02736
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids

; TYPE: amino acid
; STRANDEDNESS: NO. US20020187486A1 Relevant
; TOPOLOGY: NO. US20020187486A1 Relevant
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 143:
US-10-033-297-143

Query Match      80.3%; Score 110; DB 13; Length 543;
Best Local Similarity 78.6%; Pred. No. 1.2e-07;
Matches 22; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LKQLDAQQTQLRIDSFFRLAQKEKDKR 28
   |||:|||||:|||||:|||||:|||||:
Db 345 LKHLNAHQTLRIDSFFRLAQKEKQDAK 372

RESULT 10
US-10-290-386-143
; Sequence 143, Application US/10290386
; Publication No. US20030152971A1
; GENERAL INFORMATION:
; APPLICANT: Lyamichev, Victor
; APPLICANT: Neri, Bruce P.
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lukowiak, Andrew A.
; TITLE OF INVENTION: Methods and Compositions for Detecting Target Sequences
; FILE REFERENCE: FORS-07459
; CURRENT APPLICATION NUMBER: US/10/290,386
; PRIOR FILING DATE: 2002-11-07
; PRIOR APPLICATION NUMBER: 60/361,060
; PRIOR FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: 60/344,946
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: 09/713,601
; PRIOR FILING DATE: 2000-11-15
; PRIOR APPLICATION NUMBER: 09/381,212
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: 09/350,309
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 08/823,516
; PRIOR FILING DATE: 1997-03-24
; PRIOR APPLICATION NUMBER: 08/759,038
; PRIOR FILING DATE: 1996-12-02
; PRIOR APPLICATION NUMBER: 08/756,386
; PRIOR FILING DATE: 1996-11-26
; PRIOR APPLICATION NUMBER: 08/682,853
; PRIOR FILING DATE: 1996-07-12
; PRIOR APPLICATION NUMBER: 08/599,491
; PRIOR FILING DATE: 1996-01-24
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 143
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-290-386-143

Query Match      80.3%; Score 110; DB 14; Length 543;
Best Local Similarity 78.6%; Pred. No. 1.2e-07;
Matches 22; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LKQLDAQQTQLRIDSFFRLAQKEKDKR 28
   |||:|||||:|||||:|||||:|||||:
Db 345 LKHLNAHQTLRIDSFFRLAQKEKQDAK 372

RESULT 11
US-10-356-861-143
; Sequence 143, Application US/10356861
; Publication No. US20040072182A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Victor, Lyamichev
; APPLICANT: Neri, Bruce P.
; APPLICANT: Hall, Jeff
; APPLICANT: Lukowiak, Andrew A.
; TITLE OF INVENTION: Methods and Compositions for Detecting Target Sequences
; FILE REFERENCE: FORS-07813
; CURRENT APPLICATION NUMBER: US/10/356,861
; CURRENT FILING DATE: 2003-02-03
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 143
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-356-861-143

Query Match      80.3%; Score 110; DB 15; Length 543;
Best Local Similarity 78.6%; Pred. No. 1.2e-07;
Matches 22; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy      1 LKQLDAQOTQLRIDSPFRLAQOEKDKR 28
      |||:|||||:|||||:|||||:|:
Db      345 LKHLNAHQTLRIDSPFRLAQOEKQDAK 372

RESULT 12
US-10-897-793-143
; Sequence 143, Application US/10897793
; Publication No. US20050003432A1
; GENERAL INFORMATION:
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor I.
; Mast, Andrea L.
; Brow, Mary Ann D.
; TITLE OF INVENTION: Detection of Nucleic Acids By Multiple
; Sequential Invasive Cleavages
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/897,793
; FILING DATE: 23-Jul-2004
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/01072
; FILING DATE: 21-JAN-1997
; APPLICATION NUMBER: US 08/759,038
; FILING DATE: 02-DEC-1996
; APPLICATION NUMBER: US 08/758,314
; FILING DATE: 02-DEC-1996
; APPLICATION NUMBER: US 08/756,386
; FILING DATE: 29-NOV-1995
; APPLICATION NUMBER: US 08/682,853
; FILING DATE: 12-JUL-1996
; APPLICATION NUMBER: US 08/599,491
; FILING DATE: 24-JAN-1996
; APPLICATION NUMBER: US 08/823,516
; FILING DATE: 24-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: MacKnight, Kamrin T.
; REGISTRATION NUMBER: 38,230

```

```

; REFERENCE/DOCKET NUMBER: FORS-03295
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 143:
US-10-897-793-143

Query Match      80.3%; Score 110; DB 16; Length 543;
Best Local Similarity 78.6%; Pred. No. 1.2e-07;
Matches 22; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy      1 LKQLDAQOTQLRIDSPFRLAQOEKDKR 28
      |||:|||||:|||||:|||||:|:
Db      345 LKHLNAHQTLRIDSPFRLAQOEKQDAK 372

RESULT 13
US-10-783-557-144
; Sequence 144, Application US/10783557
; Publication No. US20050048527A1
; GENERAL INFORMATION:
; APPLICANT: Allawi, Hatim T.
; APPLICANT: Kaiser, Michael W.
; APPLICANT: Ma, Wu-Po
; APPLICANT: Neri, Bruce P.
; APPLICANT: Lyamichev, Victor I.
; TITLE OF INVENTION: Endonuclease-Substrate Complexes
; FILE REFERENCE: FORS-08907
; CURRENT APPLICATION NUMBER: US/10/783,557
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 533
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 144
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-783-557-144

Query Match      80.3%; Score 110; DB 17; Length 543;
Best Local Similarity 78.6%; Pred. No. 1.2e-07;
Matches 22; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy      1 LKQLDAQOTQLRIDSPFRLAQOEKDKR 28
      |||:|||||:|||||:|||||:|:
Db      345 LKHLNAHQTLRIDSPFRLAQOEKQDAK 372

RESULT 14
US-09-940-244-144
; Sequence 144, Application US/09940244
; Publication No. US20030044796A1
; GENERAL INFORMATION:
; APPLICANT: Neri, Bruce P.
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor
; APPLICANT: Smith, Lloyd M.
; TITLE OF INVENTION: Reactions on Dendrimers
; FILE REFERENCE: FORS-06478
; CURRENT APPLICATION NUMBER: US/09/940,244
; CURRENT FILING DATE: 2002-05-06
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 144
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Xenopus laevis

```

US-09-940-244-144

Query Match 70.1%; Score 96; DB 10; Length 527;
Best Local Similarity 87.0%; Pred. No. 1.3e-05;
Matches 20; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LKOLDAQOQTOLRIDSFRLAQOE 23
||||:|||||
Db 345 LKOLNAQOQTOLRIDSFRLAQOE 367

RESULT 15

US-10-033-297-144
; Sequence 144, Application US/10033297
; Publication No. US20020187486A1

; GENERAL INFORMATION:

APPLICANT: Hall, Jeff G.
Lyamichev, Victor I.
Mast, Andrea L.
Brow, Mary Ann D.

TITLE OF INVENTION: Detection of Nucleic Acids By Multiple
Sequential Invasive Cleavages

NUMBER OF SEQUENCES: 163

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/033,297

FILING DATE: 12-NOV-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/350,597

FILING DATE: 09-JUL-1999

APPLICATION NUMBER: US/08/823,516

FILING DATE: 24-MAR-1997

APPLICATION NUMBER: PCT/US97/01072

FILING DATE: 21-JAN-1997

APPLICATION NUMBER: US 08/759,038

FILING DATE: 02-DEC-1996

APPLICATION NUMBER: US 08/758,314

FILING DATE: 02-DEC-1996

APPLICATION NUMBER: US 08/756,386

FILING DATE: 29-NOV-1996

APPLICATION NUMBER: US 08/682,853

FILING DATE: 12-JUL-1996

APPLICATION NUMBER: US 08/599,491

FILING DATE: 24-JAN-1996

ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: FORS-02736

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 144:

SEQUENCE CHARACTERISTICS:

LENGTH: 527 amino acids

TYPE: amino acid

STRANDEDNESS: No. US20020187486A1 Relevant

TOPOLOGY: No. US20020187486A1 Relevant

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 144:

US-10-033-297-144

Query Match 70.1%; Score 96; DB 13; Length 527;
Best Local Similarity 87.0%; Pred. No. 1.3e-05;
Matches 20; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LKOLDAQOQTOLRIDSFRLAQOE 23
||||:|||||
Db 345 LKOLNAQOQTOLRIDSFRLAQOE 367

Search completed: May 20, 2005, 11:14:48

Job time : 1937 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2005, 10:25:10 ; Search time 57 Seconds
(without alignments)
251.548 Million cell updates/sec

Title: US-10-083-576-1

Perfect score: 137

Sequence: 1 LKQLDAQQTQLRIDSFRLAQOQEKEDKR 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	128	93.4	415	2 Q9GMU2	Q9gm2 macaca fasc
2	128	93.4	1186	1 XPG_HUMAN	P28715 homo sapien
3	110	80.3	1170	1 XPG_MOUSE	P35689 mus musculu
4	96	70.1	1196	1 XPG_XENLA	P14629 xenopus lae
5	96	70.1	1197	2 Q6DDY8	Q6ddy8 xenopus lae
6	60	43.8	464	2 Q8W4Q3	Q8w4q3 arabidopsis
7	60	43.8	492	2 Q9LPV2	Q9lpv2 arabidopsis
8	57	41.6	493	1 GLO3_YEAST	P38682 saccharomyc
9	56	40.9	2097	2 Q7RZF8	Q7rfz8 neurospora
10	56	40.9	2115	2 Q6MFH6	Q6mfh6 neurospora
11	54.5	39.8	368	2 Q9FX50	Q9fx50 arabidopsis
12	53.5	39.1	730	2 Q9VJB4	Q9vj84 drosophila
13	53	38.7	129	1 UMPI_SCHPO	Q74416 schizosacch
14	53	38.7	372	2 Q9MAA8	Q9maa8 arabidopsis
15	52	38.0	978	2 Q9PT35	Q9pt35 staphylococ
16	52	38.0	978	2 Q7A0K3	Q7a0k3 staphylococ
17	52	38.0	978	2 Q7A4V5	Q7a4v5 staphylococ
18	52	38.0	978	2 Q6G892	Q6g892 staphylococ
19	52	38.0	978	2 Q6GFL3	Q6gfl3 staphylococ
20	51	37.2	64	2 Q9TUX9	Q9tux9 monodelphis
21	51	37.2	66	2 Q9XT04	Q9xt04 macropus ru
22	51	37.2	189	2 Q8DAV8	Q8dav8 vibrio vuln
23	51	37.2	222	2 Q7MI29	Q7mi29 vibrio vuln
24	51	37.2	329	2 Q51004	Q51004 neisseria g
25	51	37.2	336	2 Q51063	Q51063 neisseria g
26	51	37.2	389	2 Q25554	Q25554 naegleria f
27	51	37.2	557	1 TCFA_MONDO	Q9xt06 monodelphis
28	50	36.5	223	2 Q8PL51	Q8pl51 sorghum bic
29	50	36.5	355	2 Q08536	Q08536 bacteroides
30	50	36.5	381	2 Q8PL48	Q8pl48 xanthomonas
31	50	36.5	933	2 Q9W3V2	Q9w3v2 drosophila

32	50	36.5	958	2 Q8BLI4	Q8bli4 m mus muscu
33	49.5	36.1	123	2 Q6F978	Q6f978 acinetobact
34	49.5	36.1	1046	2 Q7PZ80	Q7p280 anopheles g
35	49	35.8	246	2 Q6J544	Q6j544 dendrocalam
36	49	35.8	276	2 P95429	P95429 pseudomonas
37	49	35.8	276	2 Q9I321	Q9i321 pseudomonas
38	49	35.8	311	2 Q88B21	Q88b21 pseudomonas
39	49	35.8	325	2 Q9NKP9	Q9nkp9 leishmania
40	49	35.8	355	2 Q8KVR8	Q8kvr8 bacteroides
41	49	35.8	355	2 Q45799	Q45799 bacteroides
42	49	35.8	385	2 Q8A4G8	Q8a4g8 bacteroides
43	49	35.8	447	2 Q7SD94	Q7sd94 neurospora
44	49	35.8	471	2 Q9LRV8	Q9lr8 arabidopsis
45	49	35.8	958	2 Q9UL01	Q9ul01 homo sapien

ALIGNMENTS

RESULT 1
Q9GMU2
ID Q9GMU2 PRELIMINARY; PRT; 415 AA.
AC Q9GMU2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain parietal lobe;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047633; BABI2157.1; --
DR HSSP; O93634; 1B43.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0003697; F:single-stranded DNA binding; IEA.
DR GO; GO:0006289; P:nucleotide-excision repair; IEA.
DR InterPro; IPR008918; 5.3 exo.C.
DR InterPro; IPR000513; EXO_N_I.
DR InterPro; IPR001044; XPGC_DNA_repair.
DR InterPro; IPR006084; XPGC_Rad.
DR Pfam; PF00867; XPG_I; 1.
DR PRINTS; PR00853; XPGRADSUPER.
DR PRINTS; PR00066; XRODRMPGNTG.
DR SMART; SM00279; HhH2; 1.
DR SMART; SM00484; XPGI; 1.
DR PROSITE; PS00842; XPG_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 415 AA; 46951 MW; A861857142E340BD CRC64;

Query Match 93.4%; Score 128; DB 2; Length 415;

Best Local Similarity 92.9%; Pred. No. 1.2e-09;

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LKQLDAQQTQLRIDSFRLAQOQEKEDKR 28

Db 214 LKQLDAQQTQLRIDSFRLAQOQEKEDAK 241

RESULT 2

ID_XPG_HUMAN STANDARD; PRT; 1186 AA.
AC P28715; Q7Z2V3; Q8IZL6; Q8N1B7; Q9HD59;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)

- DT 05-JUL-2004 (Rel. 44, Last annotation update)
- DE DNA-repair protein complementing XP-G cells (Xeroderma pigmentosum
- DE group G complementing protein) (DNA excision repair protein ERCC-5).
- GN Name=ERCC5; Synonyms=XPG, XPGC;
- OS Homo sapiens (Human).
- OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
- OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
- OX NCBI_TaxID=9606;
- RN [1]
- RP SEQUENCE FROM N.A.
- RX MEDLINE=93247645; PubMed=8483504; DOI=10.1038/363182a0;
- RA Scherly D., Nospikel T., Corlet J., UCLA C., Bairoch A.,
- RA Clarkson S.G.;
- RT "Complementation of the DNA repair defect in Xeroderma pigmentosum
- RT group G cells by a human cDNA related to yeast RAD2.";
- RL Nature 363:182-185(1993).
- RN [2]
- RP SEQUENCE FROM N.A.
- RX MEDLINE=94173288; PubMed=7510366; DOI=10.1016/0921-8777(94)90080-9;
- RA Shiom T., Harada Y.-N., Saito T., Shiom T., Okuno Y., Yamaizumi M.;
- RA "An ERCC5 gene with homology to yeast RAD2 is involved in group G
- RT Xeroderma pigmentosum.";
- RL Mutat. Res. 314:167-175(1994).
- RN [3]
- RP SEQUENCE FROM N.A.
- RX MEDLINE=94019313; PubMed=8413238;
- RA Macinnes M.A., Dickson J.A., Hernandez R.R., Learmonth D., Lin G.Y.,
- RA Mudgett J.S., Park M.S., Schauer S., Reynolds R.J., Strniste G.F.,
- RA Yu J.Y.;
- RT "Human ERCC5 cDNA-cosmid complementation for excision repair and
- RT bipartite amino acid domains conserved with RAD proteins of
- RT Saccharomyces cerevisiae and Schizosaccharomyces pombe.";
- RL Mol. Cell. Biol. 13:6393-6402(1993).
- RN [4]
- RP SEQUENCE FROM N.A., AND VARIANT ASP-1104.
- RX MEDLINE=21169345; PubMed=11265544; DOI=10.1093/nar/29.7.1443;
- RA Emmert S., Schneider T.D., Khan S.G., Kraemer K.H.;
- RA "The human XPG gene: gene architecture, alternative splicing and
- RT single nucleotide polymorphisms.";
- RL Nucleic Acids Res. 29:1443-1452(2001).
- RN [5]
- RP SEQUENCE FROM N.A., AND VARIANTS VAL-254 AND ASP-1104.
- RC TISSUE=Bone marrow;
- RA Zan Q., Guo J.H., Yu L.;
- RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
- RN [6]
- RP SEQUENCE FROM N.A., AND VARIANT ASP-1104.
- RA Rieder M.J., Livingston R.J., Daniels M.R., Montoya M.A., Chung M.-W.,
- RA Miyamoto K.E., Nguyen C.P., Nguyen D.A., Foel C.L., Robertson P.D.,
- RA Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;
- RT "NIH-SNPs, environmental genome project, NIHES ES15478, Department
- RT of Genome Sciences, Seattle, WA (URL: <http://egp.gs.washington.edu>).";
- RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
- RN [7]
- RP SEQUENCE FROM N.A., AND VARIANT ASP-1104.
- RC TISSUE=Eye;
- RX MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;
- RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
- RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
- RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
- RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
- RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
- RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
- RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
- RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
- RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
- RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
- RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
- RA Whiting J., Heiton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
- RA Blakesley R.W., Touchman J.W., Green E.D., Shevchenko Y., Bouffard G.G.,
- RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
- RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smailus D.E.,
- RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
- RT "Generation and initial analysis of more than 15,000 full-length human
- RT and mouse cDNA sequences.";
- RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
- RN [8]
- RP SEQUENCE OF 1-88 FROM N.A.
- RX MEDLINE=94375034; PubMed=8088806;
- RA Samer S., Jones T.A., Corlet J., Scherly D., Sheer D., Wood R.D.,
- RA Clarkson S.G.;
- RT "The human gene for Xeroderma pigmentosum complementation group G
- RT (XPG) maps to 13q33 by fluorescence in situ hybridization.";
- RL Genomics 21:283-285(1994).
- RN [9]
- RP CHARACTERIZATION.
- RX MEDLINE=94266772; PubMed=8206890;
- RA O'Donovan A., Scherly D., Clarkson S.G., Wood R.D.;
- RT "Isolation of active recombinant XPG protein, a human DNA repair
- RT endonuclease.";
- RL J. Biol. Chem. 269:15965-15968(1994).
- RN [10]
- RP CHARACTERIZATION.
- RX MEDLINE=94376899; PubMed=8090225; DOI=10.1038/371432a0;
- RA O'Donovan A., Davies A.A., Moggs J.G., West S.C., Wood R.D.;
- RT "XPG endonuclease makes the 3' incision in human DNA nucleotide
- RT excision repair.";
- RL Nature 371:432-435(1994).
- RN [11]
- RP CHARACTERIZATION.
- RX MEDLINE=94359802; PubMed=8078765;
- RA Habraken Y., Sung P., Prakash L., Prakash S.;
- RT "Human Xeroderma pigmentosum group G gene encodes a DNA
- RT endonuclease.";
- RL Nucleic Acids Res. 22:3312-3316(1994).
- RN [12]
- RP CHARACTERIZATION, AND SUBCELLULAR LOCATION.
- RX PubMed=7651464; DOI=10.1016/0165-7992(95)90070-5;
- RA Cloud K.G., Shen B., Strniste G.F., Park M.S.;
- RT "XPG protein has a structure-specific endonuclease activity.";
- RL Mutat. Res. 347:55-60(1995).
- RN [13]
- RP INTERACTION WITH PCNA.
- RX PubMed=9305916; DOI=10.1074/jbc.272.39.24522;
- RA Gary R., Ludwig D.L., Cornelius H.L., Macinnes M.A., Park M.S.;
- RT "The DNA repair endonuclease XPG binds to proliferating cell nuclear
- RT antigen (PCNA) and shares sequence elements with the PCNA-binding
- RT regions of FEN-1 and cyclin-dependent kinase inhibitor p21.";
- RL J. Biol. Chem. 272:24522-24529(1997).
- RN [14]
- RP REVIEW.
- RX PubMed=14726017; DOI=10.1016/j.biochi.2003.10.014;
- RA Clarkson S.G.;
- RT "The XPG story.";
- RL Biochimie 85:1113-1121(2003).
- RN [15]
- RP REVIEW ON VARIANTS XP-G.
- RX MEDLINE=99374920; PubMed=10447254;
- RX DOI=10.1002/(SICI)1098-1004(1999)14:1<9::AID-HUMU2>3.3.CO;2-Y;
- RA Cleaver J.E., Thompson L.H., Richardson A.S., States J.C.;
- RT "A summary of mutations in the UV-sensitive disorders: Xeroderma
- RT pigmentosum, Cockayne syndrome, and trichothiodystrophy.";
- RL Hum. Mutat. 14:9-22(1999).
- RN [16]
- RP VARIANT XP-G VAL-792, AND VARIANT ASP-1104.
- RX MEDLINE=95038755; PubMed=7951246;
- RA Nospikel T., Clarkson S.G.;
- RT "Mutations that disable the DNA repair gene XPG in a Xeroderma
- RT pigmentosum group G patient.";
- RL Hum. Mol. Genet. 3:963-967(1994).
- RN [17]
- RP VARIANT XP-G VAL-792.
- RX MEDLINE=97250499; PubMed=9096355; DOI=10.1073/pnas.94.7.3116;
- RA Nospikel T., Lalle P., Leadon S.A., Cooper P.K., Clarkson S.G.;
- RT "A common mutational pattern in Cockayne syndrome patients from

RT Xeroderma pigmentosum group G: implications for a second XPG
 RL function.";
 RN Proc. Natl. Acad. Sci. U.S.A. 94:3116-3121(1997).
 RP [18]
 RP VARIANT XP-G/CS HIS-72.
 RX MEDLINE=21126776; PubMed=11228268;
 RA Zafeiropoulos D.I., Thorel F., Andreou A., Kleijer W.J., Raams A.,
 RA Garritsen V.H., Gombakis N., Jaspers N.G.J., Clarkson S.G.;
 RT "Xeroderma pigmentosum group G with severe neurological involvement
 RT and features of Cockayne syndrome in infancy.";
 RL Pediatr. Res. 49:407-412(2001).
 RN [19]
 RP VARIANT XP-G PRO-858.
 RX MEDLINE=21830955; PubMed=11841555;
 RA Lalle P., Noupikel T., Constantinou A., Thorel F., Clarkson S.G.;
 RA "The founding members of xeroderma pigmentosum group G produce XPG
 RT protein with severely impaired endonuclease activity.";
 RL J. Invest. Dermatol. 118:344-351(2002).
 RN [20]
 RP VARIANT XP-G THR-874.
 RX MEDLINE=22053379; PubMed=12060391;
 RA Emmert S., Sior H., Busch D.B., Batko S., Albert R.B., Coleman D.,
 RA Khan S.G., Abu-libdeh B., DiGiovanna J.J., Cunningham B.B., Lee M.M.,
 RA Crollick J., Inui H., Ueda T., Hedayati M., Grossman L., Shahnavi T.,
 RA Cleaver J.E., Kraemer K.H.;
 RT "Relationship of neurologic degeneration to genotype in three
 RT xeroderma pigmentosum group G patients.";
 RL J. Invest. Dermatol. 118:972-982(2002).
 CC -!- FUNCTION: Single-stranded structure-specific DNA endonuclease
 CC involved in DNA excision repair. Makes the 3' incision in DNA
 CC nucleotide excision repair (NER). Acts as a cofactor for a DNA
 CC glycosylase that removes oxidised pyrimidines from DNA. May also
 CC be involved in transcription-coupled repair of this kind of
 CC damage, in transcription by RNA polymerase II, and perhaps in
 CC other processes too.
 CC -!- SUBUNIT: Interacts with PCNA.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- DISEASE: Defects in ERCC5 are the cause of xeroderma pigmentosum
 CC complementation group G (XP-G) (MIM:278870); also known as
 CC xeroderma pigmentosum VII (XP7). XP-G is an autosomal recessive
 CC disease characterized by hypersensitivity of the skin to sunlight
 CC followed by high incidence of skin cancer and frequent neurologic
 CC abnormalities.
 CC -!- DISEASE: Defects in ERCC5 are one of the cause of Cockayne's
 CC syndrome (CS) (MIM:216400). CS is an autosomal recessive disease
 CC which is characterized by a UV-sensitive skin (without
 CC pigmentation abnormalities), neurological dysfunction due to
 CC demyelination of neurons and calcification of basal ganglia
 CC (psychomotor retardation, deafness, optic atrophy, retinal
 CC pigmentation and hyperreflexes), and dysmorphic dwarfism (immature
 CC Query Match 93.4%; Score 128; DB 1; Length 1186;
 CC .Best Local Similarity 92.9%; Pred. No. 3.8e-09;
 CC Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 LKQDAQQTQLRIDSFFFLAQEKEDKR 28
 Db 981 LKQDAQQTQLRIDSFFFLAQEKEDAK 1008
 RESULT 3
 ID_XPG_MOUSE STANDARD; PRT; 1170 AA.
 AC P35689; O61528; Q64248;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 25-OCT-2004 (Rel. 45, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE DNA-repair protein complementing XP-G cells homolog (Xeroderma
 DE pigmentosum group G complementing protein homolog) (DNA excision
 DE repair protein ERCC-5).
 GN Name=Ercc5; Synonyms=Ercc-5, Xpg;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94173288; PubMed=7510366; DOI=10.1016/0921-8777(94)90080-9;
 RA Shlomi T., Harada Y.-N., Saito T., Shioni N., Okuno Y., Yamazumi M.;
 RT "An ERCC5 gene with homology to yeast RAD2 is involved in group G
 RT Xeroderma pigmentosum.";
 RL Mutat. Res. 314:167-175(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=C57BL/10; TISSUE=Liver;
 RX MEDLINE=96070433; PubMed=7590748;
 RA Harada Y.N., Matsuda Y., Shioni N., Shioni T.;
 RT "Complementary DNA sequence and chromosomal localization of xpg, the
 RT mouse counterpart of human repair gene XPG/ERCC5.";
 RL Genomics 28:59-65(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=DBA/2;
 RX MEDLINE=96359149; PubMed=8703115;
 RA Ludwig D.L., Mudgett J.S., Park M.S., Perez-Castro A.V.,
 RA MacInnes M.A.;
 RT "Molecular cloning and structural analysis of the functional mouse
 RT genomic XPG gene.";
 RL Mamm. Genome 7:644-649(1996).
 CC -!- FUNCTION: Single-stranded structure-specific DNA endonuclease
 CC involved in DNA excision repair. Makes the 3' incision in DNA
 CC nucleotide excision repair (NER). Acts as a cofactor for a DNA
 CC glycosylase that removes oxidised pyrimidines from DNA. May also
 CC be involved in transcription-coupled repair of this kind of
 CC damage, in transcription by RNA polymerase II, and perhaps in
 CC other processes too (By similarity).
 CC -!- SUBUNIT: Interacts with PCNA (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- SIMILARITY: Belongs to the XPG/RAD2 endonuclease family. XPG
 CC subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch)
 CC -----
 CC EMBL; D16306; BAA03813.1; -;
 CC EMBL; U40796; AAA91039.1; -;
 CC EMBL; U40795; AAB17885.1; -;
 CC EMBL; U39892; AAB17885.1; JOINED.
 CC EMBL; U39893; AAB17885.1; JOINED.
 CC EMBL; U39894; AAB17885.1; JOINED.
 CC EMBL; U39896; AAB17885.1; JOINED.
 CC EMBL; U40073; AAB17885.1; JOINED.
 CC EMBL; U40431; AAB17885.1; JOINED.
 CC EMBL; U40432; AAB17885.1; JOINED.
 CC EMBL; U40668; AAB17885.1; JOINED.
 CC EMBL; U40669; AAB17885.1; JOINED.
 CC EMBL; U40670; AAB17885.1; JOINED.
 CC EMBL; U40792; AAB17885.1; JOINED.
 CC EMBL; U40793; AAB17885.1; JOINED.
 CC EMBL; U40794; AAB17885.1; JOINED.
 CC PIR; A57650; A57650.
 CC MGI; MGI:103582; Ercc5.
 CC InterPro; IPR008918; S_3_exo_C.
 CC InterPro; IPR00513; Exo_N_I.
 CC InterPro; IPR006086; XPG_I.
 CC InterPro; IPR006085; XPG_N.
 CC InterPro; IPR001044; XPGC_DNA_repair.
 CC InterPro; IPR006084; XPGC_Rad.
 CC Pfam; PF00867; XPG_I; 1.
 CC Pfam; PF00752; XPG_N; 1.
 CC PRINTS; PR00853; XPGRADSUPER.

```

DR PRINTS; PR00066; XRODRMPGNTG.
DR SMART; SM00279; Hh2; 1.
DR SMART; SM00484; XPG1; 1.
DR SMART; SM00485; XPGN; 1.
DR TIGRFAMS; TIGR00600; rad2; 1.
DR PROSITE; PS00841; XPG 1; 1.
DR PROSITE; PS00842; XPG 2; 1.
KW DNA repair; DNA-binding; Endonuclease; Hydrolase; Nuclear protein;
Nuclease.
FT DOMAIN 1 95 N-domain.
FT DOMAIN 752 882 I-domain.
FT DOMAIN 1049 1065 Nuclear localization signal (Potential).
FT VARIANT 388 388 R -> C (in strain DBA/2).
FT VARIANT 488 488 S -> R (in strain DBA/2).
FT VARIANT 688 688 I -> T (in strain DBA/2).
FT VARIANT 1015 1015 S -> N (in strain DBA/2).
FT VARIANT 1021 1021 M -> I (in strain DBA/2).
FT VARIANT 1121 1121 S -> P (in strain DBA/2).
FT CONFLICT 227 227 N -> M (in Ref. 1).
FT CONFLICT 249 249 N -> M (in Ref. 1).
FT CONFLICT 300 302 VMD -> MDE (in Ref. 1).
FT CONFLICT 313 313 N -> M (in Ref. 1).
FT CONFLICT 320 320 N -> M (in Ref. 1).
FT CONFLICT 399 399 N -> M (in Ref. 1).
FT CONFLICT 408 408 D -> DVQTGG (in Ref. 1).
FT CONFLICT 581 581 N -> NSASEVIGPV (in Ref. 1).
FT CONFLICT 795 795 M -> V (in Ref. 1).
FT CONFLICT 1039 1039 A -> AMEKPEL (in Ref. 1).
FT CONFLICT 1134 1134 S -> SD (in Ref. 1).
FT CONFLICT 1157 1158 KL -> RR (in Ref. 1).
SQ SEQUENCE 1170 AA; 130857 MW; 41E2FACE47167A57 CRC64;

Query Match 80.3%; Score 110; DB 1; Length 1170;
Best Local Similarity 78.6%; Pred. No. 1.3e-06;
Matches 22; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LKOLDAQOQTQLRIDSPFLRAQKEKCR 28
DB 980 LKLNHAHQTLRIDSPFLRAQKEKQDAK 1007

RESULT 4
XPG_XENLA
ID XPG_XENLA STANDARD; PRT; 1196 AA.
AC P14629;
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE DNA-repair protein complementing XP-G cells homolog (Xeroderma
DE pigmentosum group G complementing protein homolog).
GN Names=ERCC5; Synonyms=XPG, XPGC;
OC Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93247645; PubMed=8483504; DOI=10.1038/363182a0;
RA Scherly D., Nussipikel T., Corlet J., Ucla C., Bairoch A.,
RA Clarkson S.G.;
RT "Complementation of the DNA repair defect in Xeroderma pigmentosum
RT group G cells by a human cDNA related to yeast RAD2.";
RL Nature 363:182-185(1993).
CC -!- FUNCTION: Single-stranded structure-specific DNA endonuclease
CC involved in DNA excision repair. Makes the 3' incision in DNA
CC nucleotide excision repair (NER). Acts as a cofactor for a DNA
CC glycosylase that removes oxidised pyrimidines from DNA. May also
CC be involved in transcription-coupled repair of this kind of
CC damage, in transcription by RNA polymerase II, and perhaps in
CC other processes too (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Belongs to the XPG/RAD2 endonuclease family. XPG

```

```

CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X69977; CAA49597.1; -
CC PIR; S35994; S35994
CC InterPro; IPR008918; 5_3_exo_C.
CC InterPro; IPR00513; Exo_N_I.
CC InterPro; IPR006086; XPG I.
CC InterPro; IPR006085; XPG N.
CC InterPro; IPR01044; XPGC DNA repair.
CC InterPro; IPR006084; XPGC_Rad.
CC Pfam; PF00867; XPG_I; 1.
CC Pfam; PF00752; XPG_N; 1.
CC PRINTS; PR00853; XPGRADSUPER.
CC PRINTS; PR00066; XRODRMPGNTG.
CC SMART; SM00279; Hh2; 1.
CC SMART; SM00484; XPG1; 1.
CC SMART; SM00485; XPGN; 1.
CC TIGRFAMS; TIGR00600; rad2; 1.
CC PROSITE; PS00841; XPG 1; 1.
CC PROSITE; PS00842; XPG 2; 1.
KW DNA repair; DNA-binding; Endonuclease; Hydrolase; Nuclear protein;
Nuclease.
FT DOMAIN 1 95 N-domain.
FT DOMAIN 786 914 I-domain.
FT DOMAIN 936 952 Nuclear localization signal (Potential).
FT DOMAIN 1079 1095 Nuclear localization signal (Potential).
SQ SEQUENCE 1196 AA; 134206 MW; 1F1C81891A3C0623 CRC64;

Query Match 70.1%; Score 96; DB 1; Length 1196;
Best Local Similarity 87.0%; Pred. No. 0.00012;
Matches 20; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LKOLDAQOQTQLRIDSPFLRAQOE 23
DB 1014 LKQLNAQOQTQLRIDSPFLRAQOE 1036

RESULT 5
Q6DDY8
ID Q6DDY8 PRELIMINARY; PRT; 1197 AA.
AC Q6DDY8;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE XPGC protein.
DE Names=XPGC;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

```



```

RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.F.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kertman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Straube R.;
RA Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL; BC077363; AAH77363.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0003697; F:single-stranded DNA binding; IEA.
DR GO; GO:0006289; P:nucleotide-excision repair; IEA.
DR InterPro; IPR008918; S_3-exo_C.
DR InterPro; IPR000513; Exo_N_I.
DR InterPro; IPR001044; XPGC_DNA_repair.
DR InterPro; IPR006084; XPGC_Rad.
DR InterPro; IPR006086; XPG_I.
DR InterPro; IPR006085; XPG_N.
DR Pfam; PF00867; XPG_1; I.
DR Pfam; PF00752; XPG_N; 1.
DR PRINTS; PR00853; XPGRADSUPER.
DR PRINTS; PR00666; XRODRMPGNTG.
DR SMART; SM00279; Hh2; 1.
DR SMART; SM00484; XPGI; 1.
DR SMART; SM00485; XPGN; 1.
DR TIGRFAMs; TIGR00600; rad2; 1.
DR PROSITE; PS00841; XPG_1; 1.
DR PROSITE; PS00842; XPG_2; 1.
SQ SEQUENCE 1197 AA; 134331 MW; 99A9988C8CDB3711 CRC64;

* Query Match 70.1%; Score 96; DB 2; Length 1197;
* Best Local Similarity 87.0%; Pred. No. 0.00012;
Matches 20; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LKQIDAQQTQLRI--DSFFRLAQOEK 23
Db 1015 LKQLNAQQTQLRIDSFRLAQOE 1037
|||||:|||||:|||||:|||||:|||||:

RESULT 6
Q8W4Q3 Q8W4Q3 PRELIMINARY; PRT; 464 AA.
AC Q8W4Q3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE At1g12970/F13K23.18.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
Ecker J.R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Banh J., Bowser L.,
RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY062103; AALJ2979.1; -.
DR EMBL; AY124880; AAM70589.1; -.
DR HSSP; P07359; 1MOZ.
DR Pfam; PF00560; LRR_1; 7.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00369; LRR_TYP; 1.
SQ SEQUENCE 464 AA; 51647 MW; DBE6F7D9FB7C4A4A CRC64;

Query Match 43.8%; Score 60; DB 2; Length 464;
Best Local Similarity 50.0%; Pred. No. 5.2;
Matches 13; Conservative 5; Mismatches 6; Indels 2; Gaps 1;

Qy 1 LKQIDAQQTQLRI--DSFFRLAQOEK 24
Db 350 LRELDLSNNQIRVLPSFFRLEKLEK 375
|||||:|||||:|||||:|||||:|||||:

RESULT 7
Q9LPV2 Q9LPV2 PRELIMINARY; PRT; 492 AA.
AC Q9LPV2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE F13K23.23 protein.
GN Name=F13K23.23;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu S.X., Sakano H., Yu G., Lee J.M., Lenz C., Pham P., Toriumi M.,
RA Chin C., Chio J., Choi E., Chung M., Gonzalez A., Hwang B., Liu A.,
RA Vaysberg M., Altafi H., Brooks S., Buehler E., Chao Q., Conn L.,
RA Conway A.B., Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B.,
RA Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W.,
RA Ecker J.R., Federspiel N.A., Theologis A.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC012187; AAF78502.1; -.
DR PIR; F86263; F86263.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003591; LRR_TYP.
DR Pfam; PF00560; LRR_1; 7.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00369; LRR_TYP; 1.
SQ SEQUENCE 492 AA; 54959 MW; 078165E2CF2B284 CRC64;

Query Match 43.8%; Score 60; DB 2; Length 492;
Best Local Similarity 50.0%; Pred. No. 5.6;
Matches 13; Conservative 5; Mismatches 6; Indels 2; Gaps 1;

Qy 1 LKQIDAQQTQLRI--DSFFRLAQOEK 24
Db 378 LRELDLSNNQIRVLPSFFRLEKLEK 403
|||||:|||||:|||||:|||||:|||||:

```

1.

```

GN Name=B2C22.040;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX897679; CA85613.1; -
DR GO; GO:0019013; C:viral nucleocapsid; IEA.
KW Nucleocapsid
SQ SEQUENCE 2115 AA; 234322 MW; 5D44A9B0A9FE537C CRC64;

Query Match 40.9%; Score 56; DB 2; Length 2115;
Best Local Similarity 48.0%; Pred. No. 98;
Matches 12; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

Qy 2 KQIDAQQTQLRIDSFRLAQKEKD 26
Db 866 KEFDAQQYQKRIDELTMTISQKKE 890

RESULT 11
Q9FX50 PRELIMINARY; PRT; 368 AA.
ID AC Q9FX50;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE T6d4.21 protein.
GN Name=T6J4.21;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
RA Buehler E., Chao Q., Chin C., Chiou J., Choi E., Gonzalez A.,
RA Howng B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
RA Lenz C., Liu A., Liu S., Mukharsky N., Pham P., Sakano H., Shinn P.,
RA Toriumi M., Vaysberg M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC011810; AAG09549.1; -
DR InterPro; IPR010683; DUF1262.
DR Pfam; PF06880; DUF1262; 1.
SQ SEQUENCE 368 AA; 42510 MW; B9B1F6D2E43135CC CRC64;

Query Match 39.8%; Score 54.5; DB 2; Length 368;
Best Local Similarity 38.7%; Pred. No. 24;
Matches 12; Conservative 8; Mismatches 6; Indels 5; Gaps 1;

Qy 3 QIDAQQTQLRIDS-----FFRLAQQEKEDKR 28
Db 257 KLEQETNLREDSDGVVWFSLRDEKDKK 287

RESULT 12
Q9VJ84 PRELIMINARY; PRT; 730 AA.
ID AC Q9VJ84;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Cg10283-PA (Cg10283-pb).

```

```

GN ORFNames=CG10283;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer V.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bernos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhao Q., Zheng L.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhou S.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Swirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RL "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RL melanogaster euchromatic genome sequence";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RA MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Swirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RL "The transposable elements of the Drosophila melanogaster euchromatin:
RL a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RA MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,

```

RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RG FLYBASE;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003659; AAF53670.1; -.
DR IntAct; Q9VJ84; -.
DR FLYBASE; FBgn0032681; CG10283.
SQ SEQUENCE 730 AA; 81436 MW; 0C4E7587D5C612D2 CRC64;
Query Match 39.1%; Score 53.5; DB 2; Length 730;
Best Local Similarity 44.0%; Pred. No. 70;
Matches 11; Conservative 7; Mismatches 6; Indels 1; Gaps 1;
QY 5 DAQQTQLRIDSP-FRLAQKEKDKR 28
DB 75 EKEQQRVDVYRRRLTEQEEHDKR 99
RESULT 13
UMPI-SCHPO STANDARD; PRT; 129 AA.
AC 074416;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Putative proteasome maturation factor umpl.
GN ORFNames=SPCC14G10.03c;
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Pest N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moulis S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares S., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Furnelle V.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe".

RL Nature 415:871-880(2002).
CC -!- FUNCTION: Short-lived chaperone present in the precursor form of
CC the 20S proteasome and absent in the mature complex. Required for
CC the correct assembly and enzymatic activation of the proteasome
CC (By similarity).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL031518; CAA20656.1; -.
DR PIR; T41014; T41014.
DR GenedB Spombe; SPCC14G10.03c; -.
DR InterPro; IPR008012; UMP1.
DR Pfam; PF05348; UMP1; 1.
KW Chaperone; Hypothetical protein; Proteasome.
SQ SEQUENCE 129 AA; 14661 MW; 36CAE6B313BE4C1 CRC64;
Query Match 38.7%; Score 53; DB 1; Length 129;
Best Local Similarity 39.4%; Pred. No. 13;
Matches 13; Conservative 6; Mismatches 8; Indels 6; Gaps 1;
QY 1 LKQDAQQTQLRIDSPFRL-
DB 42 LKNWEAQQQQIRLDSMRRIYGLHLPVRREMEQK 74
RESULT 14
Q9MAA8 PRELIMINARY; PRT; 372 AA.
ID AC Q9MAA8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE T12H1.7 protein.
GN Name=T12H1.7;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lin X., Kaul S., Town C.D., Benito M.-I., Creasy T.H., Haas B., Wu D.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC009177; AAF27017.1; -.
SQ SEQUENCE 372 AA; 43077 MW; 024DBAFC8872F6D2 CRC64;
Query Match 38.7%; Score 53; DB 2; Length 372;
Best Local Similarity 41.9%; Pred. No. 40;
Matches 13; Conservative 10; Mismatches 4; Indels 4; Gaps 2;
QY 1 LKQDAQQTQLR-IDS-
DB 66 LKQLEARENELVEAKRKFFDLKEKELEK 96
RESULT 15
Q99T35 PRELIMINARY; PRT; 978 AA.
ID AC Q99T35;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=SAV1843;
OS Staphylococcus aureus (strain Mu50 / ATCC 700699).

OC Bacteria; Firmicutes; Bacilliales; Staphylococcus.
 OX NCBI_TaxID=158878;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MUS0 / ATCC 700699;
 RX MEDLINE=21311952; PubMed=11418146; DOI=10.1016/S0140-6736(00)04403-2;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hirakawa H., Kunara S., Goto S., Yabuzaki J., Shiba T.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 aureus".
 RL Lancet 357:1225-1240(2001).
 DR EMBL; AF003363; BAB58005.1; -.
 DR PIR; B89971; B89971.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 978 AA; 114404 MW; A41F40B086E98BC2 CRC64;
 Query Match 38.0%; Score 52; DB 2; Length 978;
 Best Local Similarity 45.5%; Pred. No. 1.6e+02;
 Matches 15; Conservative 2; Mismatches 10; Indels 6; Gaps 1;
 QY 1 LKQLDAQQTQLR-----IDSFPRLAQOEKEDK 27
 Db 295 LAQLKEATQLPVPKQSDIDAFISLNQENEIK 327

Search completed: May 20, 2005, 10:26:17
 Job time : 59 secs

THIS PAGE BLANK (USPTO)